Elliptical Anisotropy Statistics of Two-Dimensional Differentiable Gaussian Random Fields: Joint Probability Density Function and Confidence Regions

Manolis P. Petrakis and Dionissios T. Hristopulos

Abstract

Two-dimensional data often have autocovariance functions with elliptical equipotential contours, a property known as statistical anisotropy. The anisotropy parameters include the tilt of the ellipse (orientation angle) θ with respect to the coordinate system and the ratio R of the principal correlation lengths. Sample estimates of anisotropy parameters are needed for defining suitable spatial models and for interpolation of incomplete data. The sampling joint probability density $f_{\theta,R}(\hat{R},\hat{\theta})$ characterizes the distribution of anisotropy statistics $(\hat{R},\hat{\theta})$. By means of analytical calculations, we derive an explicit expression for $f_{\theta,R}(\hat{R},\hat{\theta})$, which is valid for Gaussian, stationary and differentiable random fields. Based on it, we derive an approximation $f^{(0)}_{\theta,R}(\hat{R},\hat{\theta})$ that is independent of the autocovariance function and provides conservative confidence regions for the sample-based estimates $(\hat{R},\hat{\theta})$. We also formulate a statistical test for isotropy based on the approximation $f^{(0)}_{\theta,R}(\hat{R},\hat{\theta})$. The proposed $f^{(0)}_{\theta,R}(\hat{R},\hat{\theta})$ provides (i) a stand-alone approximate estimate of the $(\hat{R},\hat{\theta})$ distribution (ii) computationally efficient initial values for maximum likelihood estimation, and (iii) a useful prior for Bayesian anisotropy inference. We validate the theoretical analysis by means of simulations, and we illustrate the use of confidence regions with a real-data case study.

Index Terms

Confidence regions, Machine learning, Monte Carlo simulations, Parameter estimation, Spatial statistics, Stochastic analysis.

I. Introduction

SPATIAL random fields (SRFs), also known as spatial random functions, are used in several scientific and engineering disciplines that study spatially distributed processes (e.g., image processing, theory of transport in heterogeneous media, wave propagation in random media, environmental modeling). SRFs with Gaussian joint probability density function are also used in machine learning, where they are known as *Gaussian processes* [1]. For convenience, isotropic SRF models are often used, even though many real data sets display anisotropic patterns. *Physical anisotropy* implies different values of a specific variable along different directions and is expressed by means of tensor coefficients (e.g., electrical conductivity tensor). *Statistical anisotropy* characterizes scalar processes (e.g., values of gray-scale images, pollutant concentrations), the correlation range of which depends on the spatial direction. Herein we focus on statistical anisotropy, which implies SRFs with autocovariance functions that possess elliptical equipotential contours. Estimation of anisotropy parameters is a topic of ongoing research activity in various signal and image processing applications [2]–[9], as well as in data assimilation.

In two spatial dimensions, the anisotropy parameters involve the orientation of the principal axes of anisotropy and the anisotropy ratio, i.e., the ratio of the principal correlation lengths. A problem of practical interest is the estimation of anisotropy parameters from a single available sample. The *Covariance Hessian Identity (CHI)*

M. P. Petrakis is with the Department of Mineral Resources Eng., Technical University of Crete, Chania 73100, Greece, (e-mail: petrakis@mred.tuc.gr).

Corresponding author, D. T. Hristopulos is with the Department of Mineral Resources Eng., Technical University of Crete, Chania 73100, Greece, (e-mail: dionisi@mred.tuc.gr).

This work was partially funded by INTAMAP, a project funded by the European Commission (DG INFSO, FP6). More information on INTAMAP can be found at www.intamap.org.

method [10], [11] is a non-parametric, non-iterative method for obtaining semi-analytic estimates[1] $(\hat{R}, \hat{\theta})$ of SRF anisotropy parameters from two-dimensional data sets. Nevertheless, anisotropy estimates are statistics, i.e. random variables, the values of which fluctuate between different samples. The fluctuations in the two parameters, as we show below, are inter-dependent and non-Gaussian. If the joint distribution of the fluctuations is known, it is possible to evaluate whether deviations of anisotropy statistics between different data sets are statistically significant.

Improved methods for estimating anisotropy are necessary in various research fields. The characterization and measurement of anisotropy in biological materials is important for diagnostic and medical reasons [8], [12]. Significant changes in anisotropy over time may suggest a crucial change in the underlying physical processes. For example, an accidental release of radioactivity may significantly alter the anisotropy of radioactivity patterns over the monitored area. Reliable and computationally fast detection of systematic changes in spatial distributions is crucial, especially for automatic monitoring systems [13]. Another practical question is whether an estimated anisotropy value actually means significant departure from isotropy; if not, simpler, isotropic autocovariance functions can be used for spatial modeling. Answering questions such as the above requires mathematical expressions for the confidence regions of anisotropy statistics. Herein we provide closed-form expressions for the confidence regions and the probability distribution of anisotropy statistics, which can be used as a prior in Bayesian inference [14].

Below, we derive a non-parametric approximation of the *sampling joint probability density function (JPDF)* of anisotropy statistics for differentiable, Gaussian random fields. We prove this expression using CHI, the Central Limit Theorem, Jacobi's multivariate transformation theorem, and perturbation expansion. The term "non-parametric" implies that the approximation is independent of the SRF autocovariance function (henceforward, *covariance function* for simplicity). The non-parametric approximation is shown to yield a sampling JPDF which is more dispersed in parameter space than the exact JPDF. This implies a wider confidence region for the anisotropy statistics. Hence, if a sample is classified as isotropic at confidence level p based on the approximate JPDF, it is actually isotropic at $p' > p$.

The remainder of this manuscript is structured as follows: In Section II we present essential definitions and an overview of CHI, on which the mathematical development of the joint PDF is based. In Section III we derive a general expression for $f_{\theta,R}(\hat{R}, \hat{\theta})$, using the Central Limit Theorem and the conservation of probability under variable transformation. In addition, we obtain an equation that determines the confidence regions of anisotropy statistics. In Section IV we derive the non-parametric approximation of $f_{\theta,R}(\hat{R}, \hat{\theta})$ and the corresponding confidence region expression. In Section V we formulate a non-parametric test for isotropy. In Section VI we compare the non-parametric estimates of confidence regions with the results of numerical simulations, and we illustrate the application of confidence regions using real data. Finally, in Section VII we review the main results obtained in this work, we present our conclusions, and we outline directions for future research.

II. Preliminaries

In the following, boldface symbols are used for vectors, matrices and tensors; the superscript "t" denotes the transpose of a vector or matrix. Let $\mathcal{D} \subset \mathbb{R}^2$ denote the spatial domain and $|\mathcal{D}|$ the enclosed area. The vector $\mathbf{s} \in \mathcal{D}$ denotes the position of a point in $\mathcal{D}$ and $\|\mathbf{s}\|$ denotes the Euclidean norm of $\mathbf{s}$. Let $X(\mathbf{s}, \omega)$ represent a scalar SRF on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ [2]. The state index ω determines the field state and will be suppressed in the following for the sake of brevity. The SRF $X(\mathbf{s})$ represents a scalar variable, e.g., temperature, dose rates, or grey-scale intensity levels of a digital image. The events in $\mathcal{F}$ comprise the measured SRF realization(s) or *sample states(s)*. $\mathbb{E}[\cdot]$ denotes the expectation over the ensemble of states, and the operator $\mathrm{Cov}(Z_1, Z_2) = \mathbb{E}[Z_1 Z_2] - \mathbb{E}[Z_1]\,\mathbb{E}[Z_2]$ denotes the covariance of the random variables Z_1 and Z_2.

We will focus on Gaussian SRFs (GSRFs) that possess normal joint probability density functions with variance σ_{x}^2. *Wide-sense stationarity* will be assumed, i.e. that the mean $m_{\mathrm{x}} = \mathbb{E}[X(\mathbf{s})]$ is constant, and the *covariance function* $c_{\mathrm{xx}}(\mathbf{r}) = \mathbb{E}[X(\mathbf{s})\,X(\mathbf{s}+\mathbf{r})] - m_{\mathrm{x}}^2$ is independent of $\mathbf{s}$. Let $\vec{\mathbf{e}}_i$ denote the unit vector in the spatial direction i, $(i = 1, 2)$. If the partial derivatives $\partial^2 c_{\mathrm{xx}}(\mathbf{r})/\partial r_i^2$ in the orthogonal directions $i = 1, 2$ exist at $\mathbf{r} = (0, 0)$, the SRF is differentiable in the mean square sense (m.s.s) for every $\mathbf{s} \in \mathcal{D}$, i.e., $\lim_{h\to 0} \mathbb{E}\left[\left|\partial_i X(\mathbf{s}) - \frac{[X(\mathbf{s}+h\vec{\mathbf{e}}_i) - X(\mathbf{s})]}{h}\right|^2\right] = 0$, $\forall i = 1, 2$. For Gaussian SRFs, mean square differentiability practically implies that the derivatives of the sample

[1] We use the hat over a mathematical symbol to denote *random variables* that are sample-based estimates (statistics).

[2] Ω denotes the sample space (ensemble) that includes all the possible states (realizations) of the SRF, $\mathcal{F}$ is the set of all observable events, $\mathcal{F} \subset \Omega$, and $P(\mathcal{F}) \in [0, 1]$ is the probability associated with each event.

states exist almost surely [15]. We will focus on SRFs that possess at least continuous first-order derivatives in the mean square sense. In addition, we will assume *short-range* correlations, namely that the covariance falls off sufficiently fast as $\|\mathbf{r}\| \to \infty$, so that the *correlation area* $V_c := \frac{1}{\sigma_{\mathrm{x}}^2}\int d\mathbf{r}\, c_{\mathrm{xx}}(\mathbf{r})$ is finite.

Furthermore, we assume that the *sample*, X_k^* comprises the values $X_k^* = X(\mathbf{s}_k, \omega^*)$ of the field $X(\mathbf{s})$ for a specific state ω^*, where $\mathbf{s}_k$, $k = 1, \ldots, N$ are sampling locations. We will denote sample-based estimates by the "hat" symbol, e.g., $\hat{R}$ represents the estimate of the anisotropic ratio R.

The *Covariance Hessian Matrix* $\mathbf{H}$ (CHM) of a stationary, at least once differentiable, SRF $X(\mathbf{s})$ is defined as follows

$$H_{ij}(\mathbf{r}) = -\frac{\partial^2 c_{\mathrm{xx}}(\mathbf{r})}{\partial r_i\, \partial r_j}, \quad i,j = 1,2. \tag{1}$$

Let $X_{ij} = \partial_i X(\mathbf{s})\, \partial_j X(\mathbf{s})$, $i = 1,2$ be the *Gradient Kronecker Product (GKP)* tensor; $\mathbf{X}$ is a symmetric second-rank tensor. The expectation of GKP, henceforward called the *mean slope tensor* and denoted by $\mathbf{Q}$, is defined as follows

$$Q_{ij} = \mathbb{E}\left[\partial_i X(\mathbf{s})\, \partial_j X(\mathbf{s})\right]. \tag{2}$$

The mean slope tensor is also known as the matrix of spectral moments, and plays a key role in determining the local maxima and excursion sets of random fields [16]. Swerling has proved the following *Covariance Hessian Identity* (CHI) [17]:

Theorem II.1 (Swerling's Covariance Hessian Identity)**.** *Let $X(\mathbf{s})$ be a statistically stationary GSRF with a covariance function that admits partial derivatives $\partial^2 c_{\mathrm{xx}}(\mathbf{r})/\partial r_i^2$ in the orthogonal directions $i = 1,2$ at $\mathbf{r} = (0,0)$. Then, the mean slope tensor is connected to the covariance Hessian matrix as follows:*

$$\mathbf{Q} = \mathbf{H}(\mathbf{r})|_{\mathbf{r}=\mathbf{0}}\,. \tag{3}$$

Based on *Jensen's inequality*, $\left(\mathbb{E}\left[\partial_1 X(\mathbf{s})\, \partial_2 X(\mathbf{s})\right]\right)^2 \le \mathbb{E}\left[\{\partial_1 X(\mathbf{s})\}^2\right]\, \mathbb{E}\left[\{\partial_2 X(\mathbf{s})\}^2\right]$ it follows that $\det(\mathbf{Q}) \ge 0$, where $\det(\mathbf{Q})$ is the determinant of the matrix $\mathbf{Q}$.

CHI is valid in any number of spatial dimensions. In two dimensions, the anisotropy parameters (R, θ) satisfy the following theorem [11]:

Theorem II.2. *Let $X(\mathbf{s})$ be a statistically stationary and anisotropic GSRF respecting the conditions of II.1, and $\xi_i, i = 1,2$ represent the principal correlation lengths of $X(\mathbf{s})$, i.e., $\xi_i^{-2} = \frac{-a}{\sigma_{\mathrm{x}}^2}\frac{\partial^2 c_{\mathrm{xx}}(\mathbf{r}')}{\partial r_i'^2}|_{\mathbf{r}'=\mathbf{0}}$ (where $a > 0$ is an $O(1)$ constant). We define the anisotropy ratio $R = \xi_2/\xi_1$, and the orientation (rotation) angle, θ, as the angle between the horizontal axis of the coordinate system and the first principal axis of the SRF (arbitrarily defined). We also define the slope tensor ratios of the elements Q_{ij}, $i,j = 1,2$:*

$$q_{\mathrm{d}} \doteq \frac{Q_{22}}{Q_{11}} = \frac{1 + R^2\tan^2\theta}{R^2 + \tan^2\theta}, \tag{4}$$

$$q_{\mathrm{o}} \doteq \frac{Q_{12}}{Q_{11}} = \frac{\tan\theta(R^2 - 1)}{R^2 + \tan^2\theta}. \tag{5}$$

Then, the anisotropy parameters are given by

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2q_{\mathrm{o}}}{1 - q_{\mathrm{d}}}\right), \tag{6}$$

$$R = \left[1 + \frac{1 - q_{\mathrm{d}}}{q_{\mathrm{d}} - (1 + q_{\mathrm{d}})\cos^2\theta}\right]^{-1/2}. \tag{7}$$

Proof: The proof, which is based on Theorem II.1, is shown in [11]. The only difference is that therein $R = R_{2(1)} = \xi_1/\xi_2$ is used, while above we defined $R = \xi_2/\xi_1$. The results from [11] apply by means of the transformation $R \to 1/R$. ∎

Equations (4)–(5) are invariant under the pair of transformations $\tan\theta \to -(\tan\theta)^{-1}$, that is, $\theta \to \theta \pm \pi/2$, and $R \to 1/R$. By restricting the parameter space to $R \in [0,\infty)$ and $\theta \in [-\pi/4, \pi/4)$, or equivalently $R \in [1,\infty)$ and $\theta \in [-\pi/2, \pi/2)$, the pair $(\hat{R}, \hat{\theta})$ satisfying (4)–(5) for given $(\hat{q}_d, \hat{q}_o)$ is unique, thus ensuring that the transformation $(\hat{q}_d, \hat{q}_o) \to (\hat{R}, \hat{\theta})$ is one-to-one. Theorem II.2 permits estimating the anisotropy parameters if the mean slope tensor can be estimated from the data [10], [11].

III. Sampling Joint PDF of Anisotropy Statistics

The sampling values of Q_{ij} will be denoted by the random variable $\hat{Q}_{ij}$. We propose the spatially averaged *slope tensor estimate* $\hat{Q}_{ij}$, $i,j=1,2$:

$$\hat{Q}_{ij} := \frac{1}{N}\sum_{k=1}^{N} X_{ij}(\mathbf{s}_k) = \frac{1}{N}\sum_{k=1}^{N} \partial_i X(\mathbf{s}_k)\,\partial_j X(\mathbf{s}_k). \tag{8}$$

For each sample of the SRF, in general a different estimate of the tensor $\hat{Q}_{ij}$ is obtained, leading through application of (6) and (7) to different estimates of $\hat{R}$ and $\hat{\theta}$. Hence, a probability distribution is obtained for $\hat{R}$ and $\hat{\theta}$. In this section we will calculate the joint PDF of $\hat{R}$ and $\hat{\theta}$. The calculation of $f_{\theta,R}(\hat{R},\hat{\theta})$ is based on the application of the classical *Central Limit Theorem* (CLT). As stated below, this implies $N \gg 1$, a condition that is satisfied in most applications of interest. The derivation of $f_{\theta,R}(\hat{R},\hat{\theta})$ involves several technical steps that include the variable transformations from $\hat{\mathbf{Q}}$ to $(\hat{q}_d,\hat{q}_o)$ and finally to $(\hat{R},\hat{\theta})$. In order to maintain focus in the main text, some of the proofs are relegated to Appendices. In Appendix A, Jacobi's theorems related to the transformation of probability distributions under a variable transformation are reviewed [18]. In Appendix B, we calculate the univariate PDFs of the gradient components and the GKP tensor elements. The covariance matrix of the GKP tensor elements is derived in Appendix C. The JPDF of the GKP tensor elements is derived in Appendix D based on the Central Limit Theorem. Appendix E presents the proof of Lemma III.3, which determines the bivariate JPDF of the average slope tensor ratios, $\hat{q}_{\mathrm{o}}$, and $\hat{q}_{\mathrm{o}}$. Finally, Appendix F proves the explicit relation for the JPDF of the anisotropy statistics, (18), as formulated in Theorem III.2.

A. Central Limit Theorem and joint PDF of

The classical CLT for scalar random variables is discussed in [19]–[21]. An extension of the classical CLT applies to vector random variables [22]:

Theorem III.1 (Multivariate CLT)**.** *Let us assume N independent and identically distributed vector variables $\mathbf{Z}_k$, $k=1,\ldots,N$ with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}_{ZZ}$. The random vector $\bar{\mathbf{Z}} = (\mathbf{Z}_1+\cdots+\mathbf{Z}_N)/N$ is asymptotically (i.e. for $N\to\infty$) normally distributed with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}_{ZZ}/N$.*

The same theorem also applies to averages of *correlated random variables*, i.e., random fields $\mathbf{Z}(\mathbf{s})$. The main requirements are stationarity and finite range of correlations. Stationarity ensures that the correlations between a pair of points only depend on the distance but not the location of the pair. If the size of the correlated areas is defined as $V_c := \max_{i,j}\left(\frac{1}{\sigma_{\mathrm{Z_i}}\sigma_{\mathrm{Z_j}}}\int d\mathbf{r}\, c_{\mathrm{Z_iZ_j}}(\mathbf{r})\right)$, the requirement is that V_c be finite and $|\mathcal{D}|/V_c \gg 1$. This follows by a straightforward extension of the scalar case [23]. An intuitive explanation of the constraint on the correlation range in the scalar case is as follows: consider that $\mathcal{D}$ comprises "blobs" of size $\propto V_c$. The point at the center of the blob is correlated with other points inside the blob and uncorrelated with points outside the blob. Hence, one can think of the blobs as representing independent random variables. Assuming a uniform distribution of sampling points, each blob contains, on average, a number of points $N_b \approx N\,V_c/\mathcal{D}$. Thus, the effective number of "independent units" is $N_{\text{eff}} = N/N_b \approx |\mathcal{D}|/V_c$. If V_c is finite and $N_{\text{eff}} \gg 1$, CLT applies to $\hat{\mathbf{q}}$ given by (8). Note that $N_{\text{eff}} \gg 1$ also implies $N \gg 1$, since $N = N_{\text{eff}} \times N_b$.

B. Joint PDF of slope tensor elements

We define the vector $\vec{\mathbb{Q}} \equiv (\hat{Q}_{11},\hat{Q}_{22},\hat{Q}_{12})^t$ that comprises the *independent components of the slope tensor statistics*. According to (8), $\hat{Q}_{ij} = \frac{1}{N}\sum_{k=1}^{N} X_{ij}(\mathbf{s}_k)$. The univariate PDFs of the $X_{ij}(\mathbf{s}_k)$, are derived in Appendix B.

Lemma III.1 (Covariance matrix $\mathbf{C}_{\vec{\mathbb{Q}}}$)**.** *For a statistically stationary, and anisotropic GSRF, the covariance matrix $\mathbf{C}_{\vec{\mathbb{Q}}}$ is defined by $C_{ij,kl} \equiv \mathrm{Cov}\left(\hat{Q}_{ij},\hat{Q}_{kl}\right)$ for $i,j,k,l=1,2$, i.e.,*

$$\mathbf{C}_{\vec{\mathbb{Q}}} = \begin{bmatrix} C_{11,11} & C_{11,22} & C_{11,12} \\ C_{22,11} & C_{22,22} & C_{22,12} \\ C_{12,11} & C_{12,22} & C_{12,12} \end{bmatrix}. \tag{9}$$

The six independent elements (upper triangular entries) of $\mathbf{C}_{\vec{\mathbb{Q}}}$ are given by the following series

$$\begin{aligned} C_{ij,kl} &= \frac{1}{N^2}\sum_{\mathbf{r}_{nm}} \left[H_{ik}(\mathbf{r}_{nm})H_{jl}(\mathbf{r}_{nm}) + H_{il}(\mathbf{r}_{nm})H_{jk}(\mathbf{r}_{nm})\right] \\ &= \frac{1}{N}\left[Q_{ik}\,Q_{jl} + Q_{il}\,Q_{jk}\right] + \frac{1}{N^2}\sum_{\mathbf{r}_{nm}\neq\mathbf{0}} \left[H_{ik}(\mathbf{r}_{nm})H_{jl}(\mathbf{r}_{nm}) + H_{il}(\mathbf{r}_{nm})H_{jk}(\mathbf{r}_{nm})\right], \end{aligned} \tag{10}$$

where $\mathbf{r}_{nm} = \mathbf{s}_n - \mathbf{s}_m$ $(n, m = 1, \ldots, N)$ is the lag vector between two locations $\mathbf{s}_n$ and $\mathbf{s}_m$.

Proof: The proof is given in Appendix C. ■

The leading term $\frac{1}{N}\left[Q_{ik}\,Q_{jl} + Q_{il}\,Q_{jk}\right]$ in (10) leads to the non-parametric approximation of $f_{\theta,R}(\hat{R},\hat{\theta})$ as shown below. The sums over $\mathbf{r}_{nm} \neq \mathbf{0}$ contribute parametric corrections that involve the covariance function. By definition, $\mathbf{C}_{\vec{\mathbb{Q}}}$ is a symmetric matrix, namely $C_{ij,kl} = C_{kl,ij}$.

Lemma III.2 (Joint PDF of $\hat{\mathbf{Q}}$)**.** *Let a statistically stationary and anisotropic GSRF $X(\mathbf{s})$ with covariance $c_{\text{xx}}(\mathbf{r})$ that is short-ranged and its spectral density satisfies $\tilde{C}_{\text{xx}}(\mathbf{k}) \underset{k\to\infty}{\sim} o(k^{-3-\epsilon})$ where $k = \|\mathbf{k}\|$ and $\epsilon > 0$. Then, the joint PDF of $\hat{Q}_{11}$, $\hat{Q}_{12}$, and $\hat{Q}_{22}$ tends to the following trivariate Gaussian as $N \to \infty, |\mathcal{D}| \to \infty$*

$$f_{\vec{\mathbb{Q}}}(\vec{\mathbb{Q}}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}) = \frac{1}{(2\pi)^{3/2}\,\det(\mathbf{C}_{\vec{\mathbb{Q}}})^{1/2}}\, e^{-\frac{1}{2}(\vec{\mathbb{Q}}-\mathbf{m}_{\mathbb{Q}})^t \mathbf{C}_{\vec{\mathbb{Q}}}{}^{-1}(\vec{\mathbb{Q}}-\mathbf{m}_{\mathbb{Q}})}, \tag{11}$$

where the ensemble mean $\mathbf{m}_{\mathbb{Q}}$ is given by

$$\mathbf{m}_{\mathbb{Q}}{}^t = \frac{1}{N}\sum_{k=1}^{N} \mathbb{E}\left[(X_{11}(\mathbf{s}_k), X_{22}(\mathbf{s}_k), X_{12}(\mathbf{s}_k))\right] = (Q_{11}, Q_{22}, Q_{12}), \tag{12}$$

and the covariance matrix $\mathbf{C}_{\vec{\mathbb{Q}}}$ is defined by (9) *and* (10)*.*

Proof: The proof, based on Theorem III.1, is presented in Appendix D. The condition $\tilde{C}_{\text{xx}}(\mathbf{k}) \underset{k\to\infty}{\sim} o(k^{-3-\epsilon})$, $\epsilon > 0$ implies that $\forall k \to \infty$, $\exists \epsilon > 0, C_\infty > 0$, such that $\tilde{C}_{\text{xx}}(\mathbf{k}) \leq C_\infty/k^{3+\epsilon}$. This condition is not particularly restrictive, since it is satisfied by most finite-range, differentiable covariance functions, including the Gaussian, Matérn with $\nu > 1$ and Spartan covariance models. ■

The normal probability plots of $\hat{Q}_{11}$, $\hat{Q}_{12}$, $\hat{Q}_{22}$ of simulated data, shown in Fig. 6, confirm the asymptotic normality of the univariate PDFs in agreement with the CLT.

C. PDF of slope tensor ratios

Albeit $\hat{\mathbf{q}}$ involves three independent elements, the anisotropic parameters are determined from two slope tensor ratios, c.f. (4)–(5). Next, we derive the JPDF of the slope tensor ratios $f_{\mathbf{q}}(\hat{\mathbf{q}}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}})$ where $\hat{\mathbf{q}} = (\hat{q}_d, \hat{q}_o)^t$ from the joint PDF of $\hat{\mathbf{Q}}$.

Lemma III.3 (PDF of slope tensor ratios)**.** *For a statistically stationary and anisotropic GSRF $X(\mathbf{s})$ with a covariance $c_{\text{xx}}(\mathbf{r})$ that satisfies the conditions of Lemma III.2, the PDF $f_{\mathbf{q}}(\hat{\mathbf{q}}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}})$ of the slope tensor ratios $\hat{q}_d, \hat{q}_o$ is given by the expression*

$$f_{\mathbf{q}}(\hat{\mathbf{q}}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}) = \frac{K}{8A^{\frac{5}{2}}}\, e^{-\frac{C}{2}} \left[\sqrt{2\pi}\, e^{\frac{B^2}{8A}}(B^2 + 4A)\, \operatorname{erfc}\left(\frac{B}{2\sqrt{2A}}\right) - 4B\sqrt{A}\right], \tag{13}$$

where $\operatorname{erfc}(\cdot)$ is the complementary error function, A and B are functions of the random vector $\hat{\mathbf{q}}$ as well as $\mathbf{C}_{\vec{\mathbb{Q}}}$ and $\mathbf{m}_{\mathbb{Q}}$, while C and K are functions of $\mathbf{C}_{\vec{\mathbb{Q}}}$ and $\mathbf{m}_{\mathbb{Q}}$ only:

$$A(\hat{\mathbf{q}}, \mathbf{C}_{\vec{\mathbb{Q}}}) = \hat{\mathbf{q}}^t\, \mathbf{C}_{\vec{\mathbb{Q}}}^{-1}\, \hat{\mathbf{q}}, \tag{14a}$$

$$B(\hat{\mathbf{q}}, \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}) = -2\, \mathbf{m}_{\mathbb{Q}}^t\, \mathbf{C}_{\vec{\mathbb{Q}}}^{-1}\, \hat{\mathbf{q}}, \tag{14b}$$

$$C(\mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}) = \mathbf{m}_{\mathbb{Q}}^t\, \mathbf{C}_{\vec{\mathbb{Q}}}^{-1}\, \mathbf{m}_{\mathbb{Q}}, \tag{14c}$$

$$K(\mathbf{C}_{\vec{\mathbb{Q}}}) = (2\pi)^{-3/2}\, [\det(\mathbf{C}_{\vec{\mathbb{Q}}})]^{-1/2}. \tag{14d}$$

Proof: The proof is given in Appendix E. ■

The joint PDF of $\mathbf{q}$ given by (13) is quite different from the Gaussian form of the joint PDF of $\vec{\mathbb{Q}}$.

1) Dimensional analysis: Since $\hat{q}_d, \hat{q}_o$ are dimensionless, so is $f_{\mathbf{q}}$; the coefficients A, B, C, K, nevertheless, have dimensions due to their dependence on $\mathbf{C}_{\vec{\mathbb{Q}}}$ and $\mathbf{m}_{\mathbb{Q}}$. In addition, (13) implicitly depends on N through $\mathbf{C}_{\vec{\mathbb{Q}}}$, c.f. (10). Below, we express (13) in terms of dimensionless coefficients and N.

First, let $c_{\mathrm{xx}}(\mathbf{r}) = \sigma_{\mathrm{x}}^2\, \rho(\mathbf{r}; \vec{p})$ where $\vec{p} = (\xi_1, R, \theta)$ and $-1 \leq \rho(\mathbf{r}; \vec{p}) \leq 1$ is the dimensionless autocorrelation function. Based on (10) and the definition of the covariance Hessian matrix (1), it follows that $\mathbf{C}_{\vec{\mathbb{Q}}}^{-1} = N^2\, \xi_1^4\, \sigma_{\mathrm{x}}^{-4}\, \mathbb{P}(\hat{\mathbf{q}}; R, \theta)$, where $\mathbb{P}(\hat{\mathbf{q}}; R, \theta)$ is a dimensionless matrix that depends on the functional form of $\rho(\mathbf{r}; \vec{p})$. Similarly, $A = N^2\, \xi_1^4\, \sigma_{\mathrm{x}}^{-4}\, \mathcal{A}(\hat{\mathbf{q}}; R, \theta)$, $B = N^2\, \xi_1^2\, \sigma_{\mathrm{x}}^{-2}\, \mathcal{B}(\hat{\mathbf{q}}; R, \theta)$, $C = N^2\, \mathcal{C}(\hat{\mathbf{q}}; R, \theta)$, and $K = N^3\, \xi^6\, \sigma_{\mathrm{x}}^{-6}\, \mathcal{K}(\hat{\mathbf{q}}; R, \theta)$, where the scalar coefficients $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{K}$ depend implicitly on $\rho(\mathbf{r})$. Based on the above scaling relations, we propose the following transformations that involve the *dimensionless* functions $\tilde{B}(\hat{\mathbf{q}}; R, \theta), \tilde{C}(\hat{\mathbf{q}}; R, \theta), \tilde{K}(\hat{\mathbf{q}}; R, \theta)$, which are independent of N, σ_{x}^2 and ξ_1:

$$B = 2N\sqrt{2A}\tilde{B}, \tag{15a}$$

$$C = 2\, N^2\, \tilde{C}, \tag{15b}$$

$$K = \sqrt{2} A^{3/2}\, \tilde{K}. \tag{15c}$$

Then, the slope-tensor-ratios PDF (13) is expressed as follows:

$$f_{\mathbf{q}}(\hat{\mathbf{q}}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}) = \tilde{K}\, e^{-\tilde{C}\, N^2} \left[\sqrt{\pi}\, e^{\tilde{B}\, N^2} (2\tilde{B}^2\, N^2 + 1)\, \mathrm{erfc}(\tilde{B}\, N) - 2\, \tilde{B}\, N \right]. \tag{16}$$

The JPDF and the resulting confidence regions depend on R, θ and the functional form of $\rho(\mathbf{r}; \vec{p})$.

2) Asymptotic PDF Limit : Since $A > 0$ and $B < 0$, $\tilde{B} < 0$ and the argument of $\mathrm{erfc}(\tilde{B}\, N)$ in (16) is negative. Therefore, for $N \to \infty$, and $x = \tilde{B}\, N$, it follows that $\mathrm{erfc}(x) \sim 2 + e^{-x^2} \left(\frac{1}{\sqrt{\pi} x} + O(x^{-2}) \right)$ [24]. Hence, to leading-order in N, (16) is approximated as follows:

$$f_{\mathbf{q}}(\hat{\mathbf{q}}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}) = 2\sqrt{\pi}\, \tilde{K} \left(2\tilde{B}^2\, N^2 + 1 \right) e^{N^2 (\tilde{B}^2 - \tilde{C})}. \tag{17}$$

Numerical comparisons show that the absolute relative error between the exact, (16), and the approximate, (17), JPDF is less than $\sim 10^{-9}$ for $N = 50$ and $\sim 10^{-6}$ even for $N = 30$. The numerical computations are based on the non-parametric approximation of $\mathbf{C}_{\vec{\mathbb{Q}}}$ derived in Section IV.

D. Joint PDF of anisotropy statistics

Theorem III.2 (Joint PDF of anisotropy statistics)**.** *For a statistically stationary and anisotropic GSRF $X(\mathbf{s})$ with a covariance $c_{\mathrm{xx}}(\mathbf{r})$ that satisfies the conditions of Lemma III.2, the JPDF of the statistics $\hat{R}$ and $\hat{\theta}$ is given by the following equation:*

$$f_{\theta,R}(\hat{R}, \hat{\theta}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}) = \frac{2\hat{R}\, |\hat{R}^2 - 1|}{\left(\hat{R}^2 \cos^2\hat{\theta} + \sin^2\hat{\theta} \right)^3}\, f_{\mathbf{q}}(\hat{\mathbf{q}}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}), \tag{18}$$

where $f_{\mathbf{q}}(\hat{\mathbf{q}}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}})$ is given by (16)*.*

Proof: The proof is given in Appendix F. ■

The function $f_{\theta,R}(\hat{R}, \hat{\theta})$ is clearly non-Gaussian; it depends on $\mathbf{m}_{\mathbb{Q}}$ and $\mathbf{C}_{\vec{\mathbb{Q}}}$ via $f_{\mathbf{q}}(\hat{\mathbf{q}}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}})$, while $\hat{\mathbf{q}}$ is expressed in terms of $(\hat{R}, \hat{\theta})$ using (4) and (5). If degrees are used, instead of radians, $f_{\theta,R}(\hat{R}, \hat{\theta})$ should be multiplied by $\pi/180$, which is the absolute value of the Jacobian determinant of the transformation from radians to degrees.

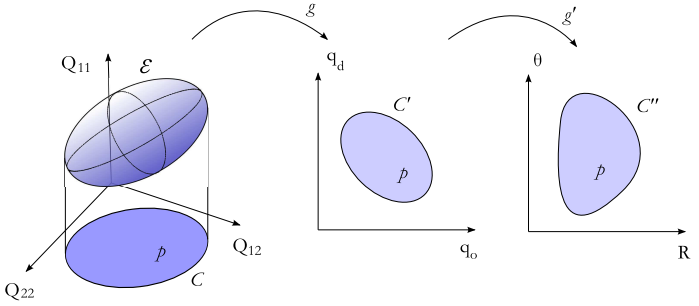


Fig. 1: Schematic illustrating the transformation of confidence regions at level p across different coordinate systems.

E. Confidence Regions

Herein, we derive expressions for confidence regions of the anisotropy statistics $(\hat{R}, \hat{\theta})$. Confidence regions are used instead of intervals due to the asymmetry of $f_{\theta,R}(\hat{R}, \hat{\theta})$. The confidence region for a probability level $p \in [0, 1]$ is the "volume" of parameter space which contains the sampled values of the statistics with probability p. The confidence region is defined by the following equivalent equations:

$$p = \begin{cases} \int_{\mathcal{E}} d\vec{\mathbb{Q}}\, f_{\vec{\mathbb{Q}}}(\hat{Q}_{11}, \hat{Q}_{22}, \hat{Q}_{12}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}), & \mathcal{E} \subset \mathbb{R}^3 \\ \int_{\mathcal{C}'} d\hat{q}_d d\hat{q}_o\, f_{\mathbf{q}}(\hat{q}_d, \hat{q}_o; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}), & \mathcal{C}' \subset \mathbb{R}^2 \\ \int_{\mathcal{C}''} d\hat{R} d\hat{\theta}\, f_{\theta,R}(\hat{R}, \hat{\theta}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}). & \mathcal{C}'' \subset [0, \infty) \times [-\pi/4, \pi/4) \end{cases}$$

The above equations represent the evolution of the confidence region under the variable transformations $\vec{\mathbb{Q}} \to \hat{\mathbf{q}} \to (\hat{R}, \hat{\theta})$ as shown schematically in Fig. 1.

Corollary III.1 (Parametric equation of confidence regions)**.** *For a statistically stationary and anisotropic GSRF $X(\mathbf{s})$ with a covariance $c_{\mathrm{xx}}(\mathbf{r})$ that satisfies the conditions of Lemma III.2, the confidence region corresponding to level $p \in [0, 1]$ in $(\hat{R}, \hat{\theta})$-space is given by the parametric equation*

$$\left[\tilde{B}^2(\hat{\mathbf{q}}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}) - \tilde{C}(\hat{\mathbf{q}}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}})\right] N^2 = \ln(1 - p), \tag{19}$$

where $\tilde{B}$, $\tilde{C}$ are given respectively by (15a)–(15b) *and $\hat{\mathbf{q}}$ translates into $(\hat{R}, \hat{\theta})$ by means of* (4) *and* (5)*.*

Proof: The JPDF $f_{\vec{\mathbb{Q}}}$ is given by the trivariate Gaussian (11). Hence, the confidence region of $\vec{\mathbb{Q}}$ is an ellipsoid whose surface satisfies the equation

$$(\vec{\mathbb{Q}} - \mathbf{m}_{\mathbb{Q}})^t\, \mathbf{C}_{\vec{\mathbb{Q}}}^{-1}\, (\vec{\mathbb{Q}} - \mathbf{m}_{\mathbb{Q}}) = \ell_p, \tag{20}$$

where $\ell_p = F^{-1}(\chi^2 = p, 2)$ is the *inverse of the chi-square cumulative distribution function* with $\nu = 2$ degrees of freedom [25]. Under the transformation $\vec{\mathbb{Q}} \to \hat{\mathbf{q}}$, the ellipsoid is projected onto an ellipse, and following the transformation $\hat{\mathbf{q}} \to (\hat{R}, \hat{\theta})$ into an asymmetric convex curve, as shown schematically in Fig. 1. Based on (E-46), the equation of the the corresponding ellipsoid in $(u, \hat{q}_d, \hat{q}_o)$-space is given by

$$A(\hat{\mathbf{q}}, \mathbf{C}_{\vec{\mathbb{Q}}})\, u^2 + B(\hat{\mathbf{q}}, \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}})\, u + C(\mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}) - \ell_p = 0,$$

where the coefficients A, B, C are given by (14). For the quadratic equation to have a unique real solution $u = \hat{Q}_{11}$ for any $\hat{\mathbf{q}}$, the discriminant should vanish, i.e.

$$B^2(\hat{\mathbf{q}}; \mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}) - 4A(\hat{\mathbf{q}}; \mathbf{C}_{\vec{\mathbb{Q}}})\, \left[C(\mathbf{m}_{\mathbb{Q}}, \mathbf{C}_{\vec{\mathbb{Q}}}) - \ell_p\right] = 0. \tag{21}$$

We can verify, using (14), that (21) represents an ellipse in the space of $\hat{\mathbf{q}}$, i.e.

$$\hat{\mathbf{q}}^t \left[(\mathbf{C}_{\vec{\mathbb{Q}}}^{-1}\, \mathbf{m}_{\mathbb{Q}})\, (\mathbf{C}_{\vec{\mathbb{Q}}}^{-1}\, \mathbf{m}_{\mathbb{Q}})^t - \left(\mathbf{m}_{\mathbb{Q}}^t\, \mathbf{C}_{\vec{\mathbb{Q}}}^{-1}\, \mathbf{m}_{\mathbb{Q}} - \ell_p\right) \mathbf{C}_{\vec{\mathbb{Q}}}^{-1}\right]\, \hat{\mathbf{q}} = 0.$$

Using the dimensionless scaling functions (15) in (21) we obtain the parametric equation $2(\tilde{C} - \tilde{B}^2)N^2 = \ell_p$, where by definition $F(\ell_p, \nu = 2) = p$; since $F(x, \nu = 2) = 1 - \exp(-x/2)$ it follows that $\ell_p = -2\ln(1-p)$ finally leading to (19). ∎

IV. Non-parametric JPDF and Confidence Region

The expressions for the JPDF, $f_{\theta,R}(\hat{R},\hat{\theta};\mathbf{m}_{\mathbb{Q}},\mathbf{C}_{\vec{\mathbb{Q}}})$, and the confidence regions of the anisotropy statistics derived above depend on the matrix $\mathbf{C}_{\vec{\mathbb{Q}}}$, given by (10). This matrix involves a series that does not, in general, admit a closed form. An approximate, analytical expression can be derived by keeping only the leading term in the covariance matrix series (10). The truncation is justified if $c_{\rm xx}(\mathbf{r})$ has short-range correlations, which implies that $H_{ij}(\mathbf{r})$ decays fast for $\|\mathbf{r}\| \gg \max(\xi_1,\xi_2)$. Then, the *leading-order approximation of* $\mathbf{C}_{\vec{\mathbb{Q}}}$ is given by

$$\mathbf{C}_{\vec{\mathbb{Q}}} \approx \frac{2}{N}\begin{bmatrix} Q_{11}^2 & Q_{12}^2 & Q_{11}Q_{12} \\ Q_{12}^2 & Q_{22}^2 & Q_{12}Q_{22} \\ Q_{11}Q_{12} & Q_{12}Q_{22} & \frac{1}{2}(Q_{12}^2+Q_{11}Q_{22}) \end{bmatrix} = \mathbf{C}_{\vec{\mathbb{Q}}}^{(0)}. \tag{22}$$

Since $\mathbf{C}_{\vec{\mathbb{Q}}}^{(0)}$ does not account for the impact of correlations, we expect that it will lead to a joint PDF with higher uncertainty, and hence wider confidence regions, than the true PDF. We confirmed this hypothesis by means of numerical simulations (see Section VI-A). In addition, $\mathbf{C}_{\vec{\mathbb{Q}}}^{(0)}$, as evidenced in (10), has an $1/N$ scaling prefactor in contrast with $1/N^2$ for the truncated terms. To accommodate this sample-size dependence, the scaling relations (15) are accordingly modified below.

Theorem IV.1 (Non-parametric JPDF)**.** *For a statistically stationary and anisotropic GSRF $X(\mathbf{s})$ with covariance $c_{\rm xx}(\mathbf{r})$ that satisfies the conditions of Lemma III.2, the non-parametric approximation of the JPDF for the anisotropy statistics is expressed as*

$$f_{\theta,R}^{(0)}(\hat{R},\hat{\theta};R,\theta,N) = |\det(\mathbf{J}_{\theta,R})|\, f_{\mathbf{q}}^{(0)}(\hat{R},\hat{\theta};R,\theta,N), \tag{23a}$$

where

$$f_{\mathbf{q}}^{(0)}(\hat{R},\hat{\theta};R,\theta,N) = 2\sqrt{\pi}\,\tilde{K}_0\,\left(2\tilde{B}_0^2\,N+1\right)\,e^{N\left(\tilde{B}_0^2-1/2\right)}. \tag{23b}$$

The coefficients $\tilde{B}_0,\tilde{K}_0$ are given by the following expressions

$$\tilde{B}_0 = (2\tilde{A}_0)^{-1/2}\left[(R^2-1)(\hat{R}^2-1)\cos(2(\theta-\hat{\theta})) - (R^2+1)(\hat{R}^2+1)\right], \tag{24a}$$

$$\tilde{K}_0 = (\pi\tilde{A}_0)^{-3/2}R^3\left[(\hat{R}^2+1)-(\hat{R}^2-1)\cos(2\hat{\theta})\right]^3, \tag{24b}$$

and $\tilde{A}_0$ is

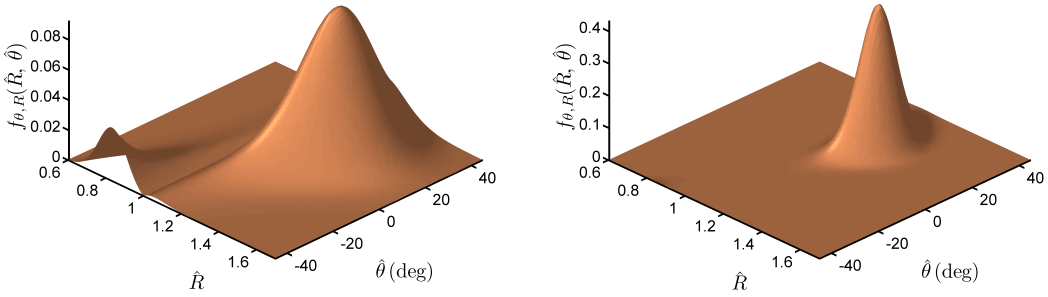
$$\begin{aligned}\tilde{A}_0 = {}& (\hat{R}^2-1)^2(R^2-1)^2\cos(4(\hat{\theta}-\theta)) - 4(\hat{R}^4-1)(R^4-1)\cos(2(\hat{\theta}-\theta)) \\ & + (\hat{R}^4+1)(3R^4+2R^2+3) + 2\hat{R}^2(R^2-1)^2.\end{aligned} \tag{24c}$$

Proof: Equations (24a)-(24c) are derived by replacing $\mathbf{C}_{\vec{\mathbb{Q}}}$ with $\mathbf{C}_{\vec{\mathbb{Q}}}^{(0)}$, defined by (22), in (14). Also. the asymptotic result (17) of Lemma III.3 was used. The $\tilde{B}_0$ and $\tilde{K}_0$ are obtained from the following dimensionless scaling functions (15a)-(15c)

$$\begin{aligned} B &= 2\sqrt{N}\,\sqrt{2A}\,\tilde{B}_0, \\ C &= 2\,N\,\tilde{C}_0, \\ K &= \sqrt{2}A^{3/2}\tilde{K}_0. \end{aligned}$$

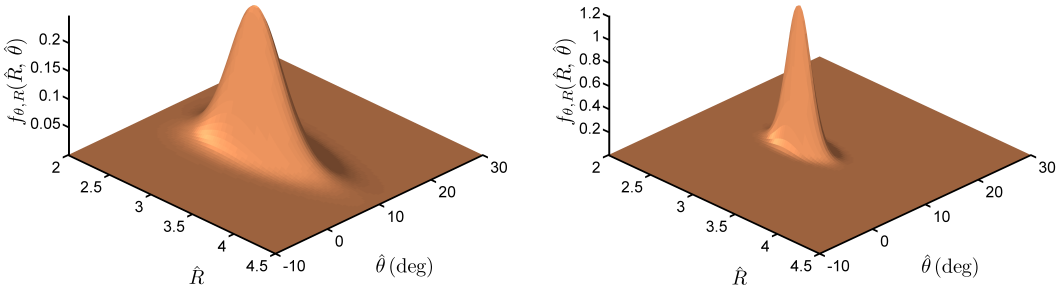
In the non-parametric approximation, the coefficient $\tilde{C}_0$ in the exponent is reduced to $1/2$ in (23b) . ■

Figures 2 and 3 demonstrate representative plots of the non-parametric JPDF based on (23). Note the bimodal structure of the JPDF for $N=100$ in Fig. 2 with one mode at $R=1.2$ and the other (smaller) at $R\approx 0.8$. This is due to the considerable spread of $\hat{\theta}$, which results from the relatively small number of sampling points, and the degeneracy of the solution, i.e., the fact that the combination (R,θ) is equivalent to $(1/R,\theta-\pi/2)$; hence, there is



(a) $N = 100$ (b) $N = 500$

Fig. 2: Non-parametric JPDF for $R = 1.2$, $\theta = 20^\circ$ and $N = 100, 500$ based on (23).



(a) $N = 100$ (b) $N = 500$

Fig. 3: Non-parametric JPDF for $R = 3$, $\theta = 10^\circ$ and $N = 100, 500$ based on (23).

an extended degenerate peak at $(0.83, -70^\circ)$, part of which is folded back into the primary domain. On the other hand, for $R = 3$ the smaller dispersion of $\hat{\theta}$ leads to a single mode even for $N = 100$.

Corollary IV.1 (Non-parametric confidence region)**.** *For a statistically stationary and anisotropic GSRF $X(\mathbf{s})$ with a covariance $c_{\mathrm{xx}}(\mathbf{r})$ that satisfies the conditions of Lemma III.2, the non-parametric equation of the confidence region corresponding to level p is given by*

$$\tilde{B}_0^2 - \frac{1}{2} = \frac{\ln(1-p)}{N}. \tag{25}$$

Proof: The equation is derived by applying the modified scaling relations of Theorem (IV.1) to the general result (21) of Theorem (III.1). ■

V. Statistical test of isotropy

Theorem V.1 (Isotropic ratio)**.** *Let $X(\mathbf{s})$ be a statistically isotropic GSRF $(R = 1)$ with correlation length ξ sampled at N points. In addition, assume that (i) $|\mathcal{D}| \gg \xi^2$ and (ii) $N \gg 1$. The sample values of the statistic $\hat{R}$ are contained with probability p in the following interval:*

$$\hat{R} \in \left[\sqrt{\frac{1 - 2\sqrt{\alpha_{p;N}(1-\alpha_{p;N})}}{1-2\alpha_{p;N}}}, \sqrt{\frac{1 + 2\sqrt{\alpha_{p;N}(1-\alpha_{p;N})}}{1-2\alpha_{p;N}}} \right], \tag{26}$$

where $\ell_p = F^{-1}(\chi^2 = p, 2) = -2\ln(1-p)$ is the inverse of the chi square cumulative distribution function with two degrees of freedom and $\alpha_{p;N} = \ell_p/N$.

Proof: $|\mathcal{D}| \gg \xi^2$ and $N \gg 1$ enforce the asymptotic conditions of Corollary IV.1. For $R = 1$ the trigonometric terms in $\tilde{B}_0$, i.e., in (24a) vanish, showing explicitly that the confidence region is independent of $\hat{\theta}$. Plugging the resulting (24a) in (25), the following parametric equation is obtained: $N(\hat{R}^2-1)^2 - 2\ell_p(\hat{R}^4+1) = 0$. The constraint

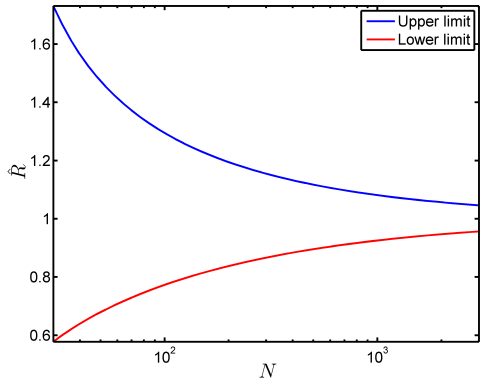


Fig. 4: Evolution of lower and upper limits of 95% confidence interval for isotropic case $(R = 1)$ versus number of sampling points N based on (26).

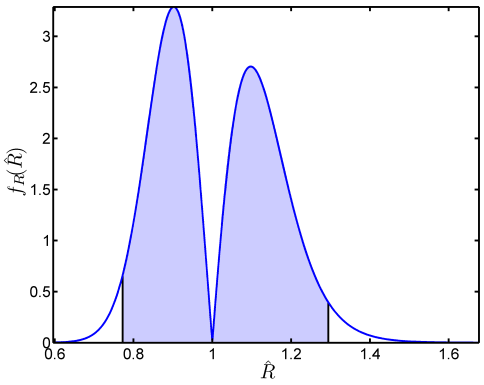


Fig. 5: Non-parametric PDF, $f_R(\hat{R})$, for isotropic case $(R = 1)$ and $N = 100$ sampling points based on (23). Shaded area represents the 95% confidence interval, $(0.77, 1.29)$, of $\hat{R}$ based on (26).

$N > 2\ell_p$ ensures that the roots are real numbers. This is satisfied for $N \gg 1$; for example, for $p = 0.95$ it implies $N > 12$. The two admissible (positive) roots of the quartic equation are given by (26). ■

Equation (26) is independent of the covariance function and thus provides a non-parametric approximation of the confidence region for $\hat{R}$. The JPDF (23) is independent of θ and $\hat{\theta}$ for $R = 1$. The dependence on N of the 95% confidence confidence interval for $\hat{R}$ is shown in Fig. 4. The PDF, $f_R(\hat{R})$, of $\hat{R}$ for $R = 1$ and $N = 100$ is shown in Fig. 5, including the 95% confidence region predicted by (26). Note that the PDF vanishes, instead of peaking, at $\hat{R} = 1$. This is not an artifact of the non-parametric approximation, since the complete PDF (18) also vanishes at $\hat{R} = 1$. This is due to the root of the Jacobian (F-47) at $\hat{R} = 1$, which reflects the fact that the single point $(1, 0)$ in $(\hat{q}_d, \hat{q}_o)$-space is mapped to the straight line $\hat{R} = 1$ in the $(\hat{R}, \hat{\theta})$-space. The vanishing of the density at $\hat{R} = 1$ is also borne out in numerical simulations that do not use the Jacobian (see Figure 8d below). This counter-intuitive behavior of $f_R(\hat{R})$, emphasizes the usefulness of Theorem V.1 as a test for isotropy.

VI. Application to Simulated and Real Data

In the following, values of statistics derived from a sample of $N \gg 1$ points will be denoted by a star superscript, e.g. $\hat{Q}^*_{ij}, \hat{R}^*, \hat{\theta}^*$. Average values of the statistics over M different samples will be denoted by a bar over the respective symbol, i.e., $\overline{Q}_{ij}$. Estimation of $\hat{Q}^*_{ij}$ is based on discretized partial derivative operators, $\hat{\partial} X_i(\mathbf{s}_k)$, where $i = 1, 2$. Discretization introduces errors that increase with the sparsity of the sampling pattern. A "good" sampling pattern is characterized by a typical distance $\hat{a}$ between nearest neighbors which is approximately uniform (ideally, a regular lattice pattern is best) and $\hat{a} \ll \min(\xi_1, \xi_2)$, where ξ_1, ξ_2 are the principal correlation lengths. Different

possibilities for $\hat{\partial} X_i(\mathbf{s}_k)$ are investigated in [11]. The parameters $\mathbf{m}_{\mathbb{Q}}$ and $\mathbf{C}_{\vec{\mathbb{Q}}}$ in (18) are unknown since they represent ensemble properties. For simulated data $\mathbf{m}_{\mathbb{Q}}$ and $\mathbf{C}_{\vec{\mathbb{Q}}}$ are replaced by the averages of the sample estimates, $\mathbf{m}_{\mathbb{Q}} \approx (\overline{Q}_{11}, \overline{Q}_{22}, \overline{Q}_{12})^t$ and $\mathbf{C}_{\vec{\mathbb{Q}}} \approx \mathbf{C}_{\overline{\mathbf{Q}}}$. In the non-parametric approximation, $\mathbf{C}^{(0)}_{\vec{\mathbb{Q}}}$ is obtained from (22) by replacing Q_{ij} with $\overline{Q}_{ij}$.

A. Simulated lattice random fields

We employ multiple SRF realizations with specified (R, θ) to validate the expression for the confidence region of the anisotropy statistics (25). The simulations are conducted on 100×100 square grids with lattice constant $a = 1$, by means of the Fourier Filtering Method [26], [27]. We use Gaussian, $c_{\rm xx}(\mathbf{r}) = \sigma_{\rm x}^2 \exp(-r^2/\xi^2)$, and Matérn, $c_{\rm xx}(\mathbf{r}) = \sigma_{\rm x}^2 \frac{2^{1-\nu} r^\nu}{\Gamma(\nu)\xi^\nu} K_\nu(r/\xi)$, covariance functions[3]. In the Gaussian case, the range of correlations is controlled by ξ but in Matérn case by both ξ and the *smoothness parameter* ν; the latter adjusts the differentiability of the random field between two extremes obtained for $\nu = 1$ (non-differentiable exponential function) and $\nu \to \infty$ (infinitely differentiable Gaussian function). For given ξ, the field becomes smoother by increasing ν. To compensate for this effect, we use *rescaled correlation lengths* $\tilde{\xi} = \xi/\ell_c$ where ℓ_c is the integral scale factor. For Matérn correlations in $d = 2$ it follows that $\ell_c = 2\sqrt{\pi\nu}$, while for Gaussian correlations $\ell_c = \pi$ [28].

The non-parametric confidence region obtained by (25), is compared with the CHI anisotropy estimates obtained from 1000 SRF samples. For each sample, we estimate the expectation (2) by means of the spatial average (8). Then, we obtain estimates $(\hat{R}^*, \hat{\theta}^*)$ of the anisotropy parameters by applying (6)–(7). We assume that (R, θ) are unknown *a priori*, and we estimate them based on $\overline{R}$, $\overline{\theta}$. The latter are obtained by calculating the average slope tensor, $\overline{\mathbf{Q}}$ over the available realizations and then applying Theorem II.2. The use of $\overline{\mathbf{Q}}$ helps to reduce biases due to finite grid size and the approximation of derivatives by means of finite differences [11]. In addition, it compensates for potential deviations of the simulated SRF from the target anisotropy values due to the Fourier filtering method.

In Figs. 6–7 we compare the cloud of $(\hat{R}^*, \hat{\theta}^*)$ estimates obtained from each realization with the non-parametric confidence region obtained by (25). The latter is denoted by the solid lines that contain the cloud. The estimated anisotropy vector, based on $\overline{\mathbf{Q}}$, is denoted by a small circle inside the cloud. In addition, we include the parametric confidence region given by (19); $\mathbf{C}_{\vec{\mathbb{Q}}}$ is estimated by numerical summation of the series (10) using a square window function of side $\sim 3\tilde{\xi}$ around a grid point $\mathbf{s}_n$ at the center of the grid. This restricts the sum over $\mathbf{s}_m$ to correlated neighbors. The summation over all $\mathbf{s}_n$ is approximated by multiplying the result with the grid size[4]. So long as (i) the window area exceeds $[2\max(\tilde{\xi}_1, \tilde{\xi}_2)]^2$ and (ii) $a < \min(\tilde{\xi}_1, \tilde{\xi}_2)$, the truncated approximation is stable for different box sizes and shapes (e.g., rectangular window). Fig. 6 investigates the *isotropic* case $(R = 1)$ while the anisotropic case $R = 1.5, \theta = -30^\circ$ is considered in Fig. 7. The non-parametric confidence region is more extended than the cloud and also encloses the parametric confidence region (contour lines inside the cloud). This result is justified, since the non-parametric approximation excludes higher-order correlated terms that reduce the uncertainty in $\mathbf{C}_{\vec{\mathbb{Q}}}$. The assumption of slope tensor normality, which was based on the application of CLT in Lemma III.2, is graphically confirmed by the normal probability plots in Figs. 6e–6g.

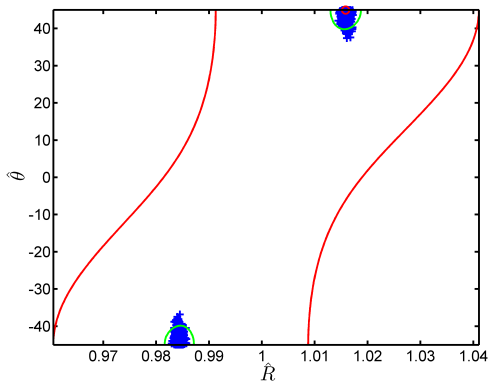
We conducted numerical experiments (not shown here) for several values of the ratio $\tilde{\xi}/a$ and various N in order to confirm that the non-parametric JPDF is more extended in parameter space more than the actual JPDF. When $\tilde{\xi}/a \to 0$, i.e., as the simulated SRF tends to random uncorrelated noise, the scatter cloud of the anisotropy estimates expands and tends to fill the non-parametric confidence region. On the other hand, as $\tilde{\xi}/a$ increases, i.e., for dense sampling of the SRF, the scatter cloud shrinks inside the corrected confidence region. These observations confirm that the non-parametric approximation encloses the actual confidence region. This is expected from an information-theoretic viewpoint, since the additional information incorporated in the covariance function should lead to less uncertainty (i.e., a tighter confidence region) than the non-parametric approximation which discards covariance terms for non-zero distances.

B. Simulated scattered data

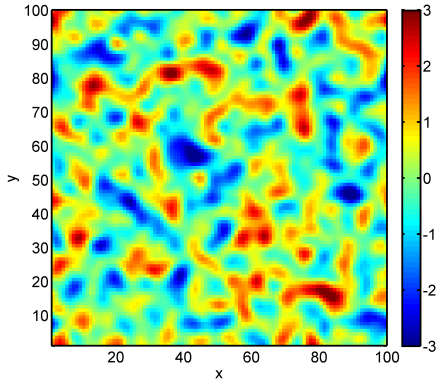
We simulate scattered data using the following method: First, a realization of an GSRF is generated on a regular grid. Then, we randomly choose a fraction of the grid points to mimic scattered data of a real process. Next,

[3]These expressions correspond to the isotropic case.

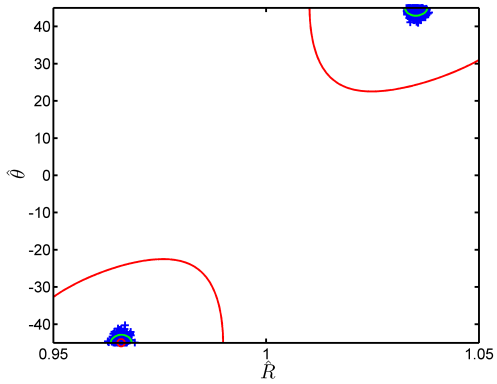
[4]This approximation introduces errors in the summands from points near the grid boundaries; however, the fraction of boundary points for an $L \times L$ grid varies as $O(1/L)$ and is thus negligible for large grids.



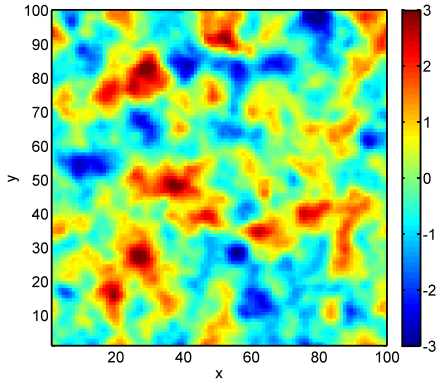
(a) Scatter plot and CR (Gaussian covariance).



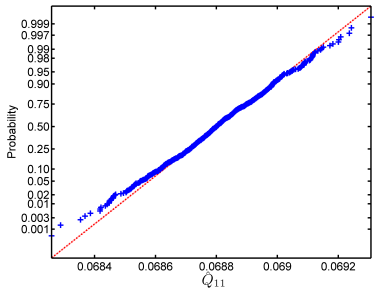
(b) Random field (Gaussian covariance).



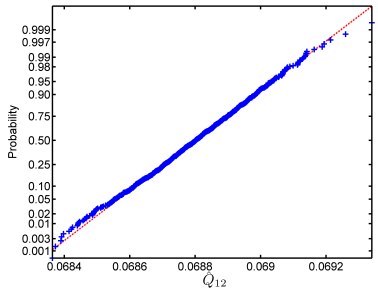
(c) Scatter plot and CR (Matérn covariance).



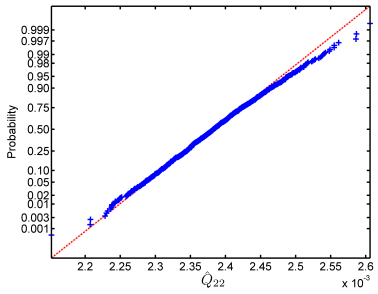
(d) Random field (Matérn covariance).



(e) $\hat{Q}_{11}$

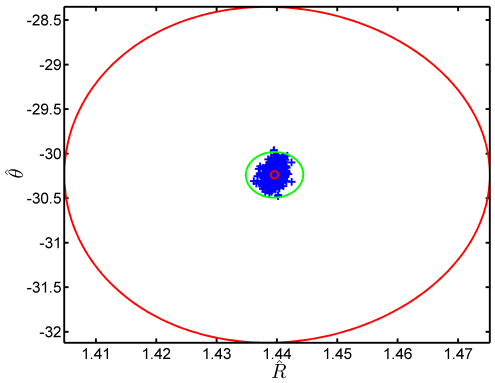


(f) $\hat{Q}_{12}$

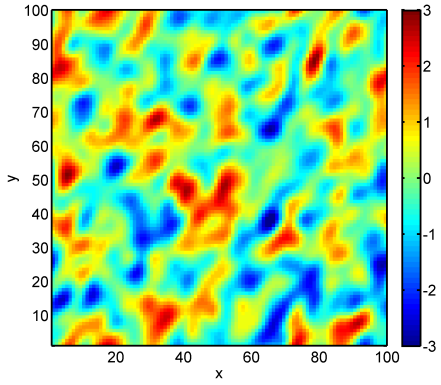


(g) $\hat{Q}_{22}$

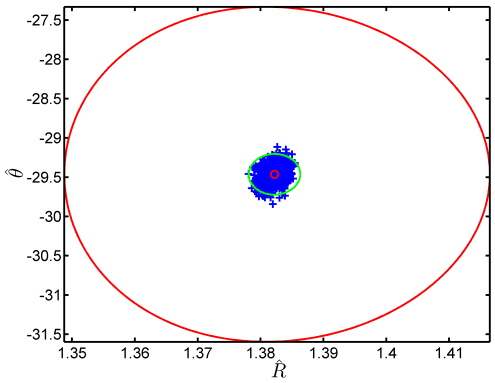
Fig. 6: (a) and (c): Scatter plot of anisotropy estimates (crosses) and non-parametric confidence region (outer contours) for 1000 realizations of isotropic random field with Gaussian $(\tilde{\xi} = 4)$ and Matérn $(\tilde{\xi} = 2.51,\ \nu = 2)$ correlations over an 100×100 square grid. A single realization for each field is shown in (b) and (d). Confidence regions (CR) are based on anisotropy parameters obtained from the mean slope tensor elements (indicated by small circle). Non-parametric confidence regions (outer contours) are obtained by (25). Tighter confidence regions (inner contours) are calculated numerically using the appropriate covariance function and incorporate the covariance-dependent terms in (10). Normal probability plots (e)–(g) for the slope tensor estimates (Matérn case) justify use of CLT in Lemma III.2.



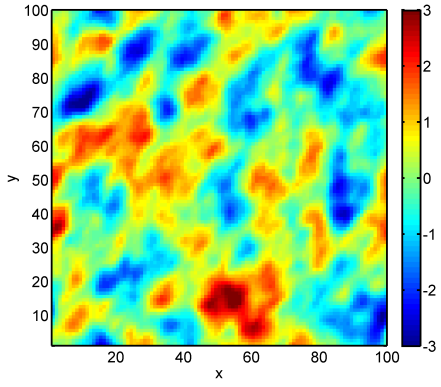
(a) Scatter plot and CR (Gaussian covariance).



(b) Random field (Gaussian covariance).



(c) Scatter plot and CR (Matérn covariance).



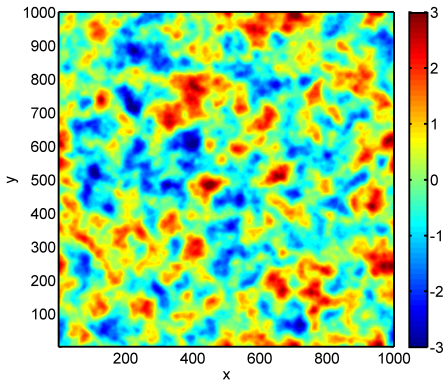
(d) Random field (Matérn covariance).

Fig. 7: (a) and (c): Scatter plot of anisotropy estimates and non-parametric confidence region (outer contour) for 1000 realizations of anisotropic random field with Gaussian ($\tilde{\xi} = 4$) and Matérn ($\tilde{\xi} = 2.51, \nu = 2$) correlations, both having $R = 1.5$, $\theta = -30°$. A single realization for each field is shown in (b) and (d). Tighter confidence region (inner contour) is numerically calculated and incorporates the covariance-dependent terms in (10).

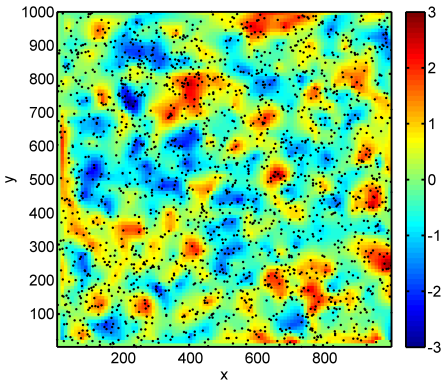
we generate several random subsamples from the scattered data set and perform anisotropy estimation for each subsample. The subsamples should respect the conditions specified in the opening paragraph of Section VI. The procedure is depicted in Figure 8. Figure 8a demonstrates an GSRF realization on a 1000×1000 grid, generated by means of the Fourier Filtering Method. The zero-mean, unit-variance isotropic GSRF has a Matérn covariance with $\nu = 2$ and $\xi = 15$. A randomly extracted set of scattered 2000 points is shown in Figure 8b. Figure 8c demonstrates a specific subsample of the scattered data set, containing 1000 points. The depicted field corresponds to the interpolation of the 1000 points used to estimate $(\hat{R}^*, \hat{\theta}^*)$. For the interpolation, we employed the natural neighbor interpolation method [29] of MATLAB®. Details of the impact of different interpolators on anisotropy estimation can be found in [11]. Figure 8d demonstrates the non-parametric confidence region (contour lines) along with the scatter cloud of anisotropy estimates, generated by randomly selecting 1000 subsamples of 1000 points from the scattered data set.

C. Environmental emergency scenario

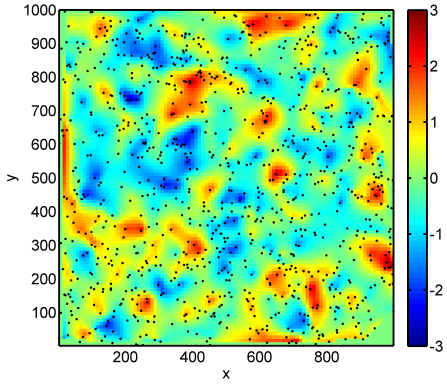
An application on an environmental emergency scenario data follows. The data represent daily means of radioactivity gamma dose rates over part of the Federal Republic of Germany, and they were provided by the German automatic radioactivity monitoring network [30]. The rates are measured in nanoSievert per hour (nSv/h). The *normal* data set corresponds to typical background radioactivity measurements (≈ 100 nSv/h). The *emergency* data includes a simulated local release of radioactivity from the south-west corner of the monitored area that results in five stations reporting dose rates around 10 times above the background (exceeding 1000 nSv/h). Table I summarizes



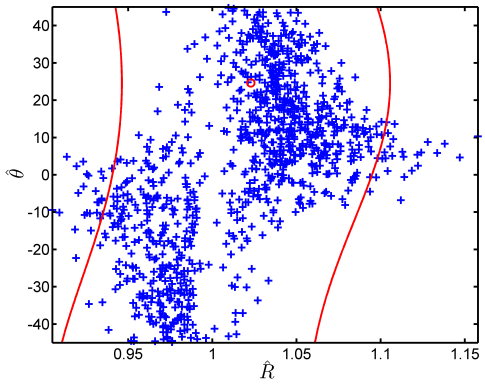
(a) Realization of Matérn SRF on 1000×1000 square grid.



(b) Scattered data locations (2000 randomly selected points) and interpolated field (for visualization purposes).



(c) Subsample realization: 1000 randomly selected locations (from those in Fig. 8b) and interpolated field (used for anisotropy estimation).



(d) Anisotropy estimates via subsampling and non-parametric CR for $N = 1000$.

Fig. 8: Non-parametric confidence region estimation for scattered data. (a) Realization of zero-mean, unit-variance isotropic Matérn SRF with $\xi = 15$ and $\nu = 2$ on a 1000×1000 square grid. (b) Locations of random sample that includes 2000 points and field realization obtained by means of natural neighbor interpolation. (c) Specific subsample realization and interpolated field used for anisotropy estimation. (d) Anisotropy estimates (crosses) generated from 1000 random subsamples of 1000 points; continuous curve corresponds to 95% non-parametric confidence region (CR) calculated with $N = 1000$ and anisotropy parameters estimated from the mean slope tensor elements as in subsection VI-A. Confidence interval for isotropy is $(\hat{R}_-, \hat{R}_+) = (0.925, 1.081)$.

TABLE I: Summary statistics of radioactivity dose rate exhaustive data sets (units are in nanoSievert per hour) and CHI anisotropy estimates. Abbreviations: min: minimum sample value; max: maximum sample value; std: sample standard deviation; sample skewness: sample skewness coefficient; kurtosis: sample excess kurtosis coefficient.

$N = 1008$	min	mean	median	max	std. dev.	skewness	kurtosis	$\hat{R}^*$	$\hat{\theta}^*(deg)$
Normal	57.0	97.7	98.6	180.0	19.6	0.4	0.6	1.18	7.36
Emergency	57.0	106.1	98.9	1528.2	92.5	11.3	144.1	0.45	-0.75

the statistics of both data sets. The two rightmost columns show the anisotropy parameters estimated with the CHI method. The normal set data follows the Gaussian distribution (graph not shown here), and thus has skewness and excess kurtosis coefficients close to zero.

We calculated the JPDF and the 95% confidence regions of the anisotropy statistics based on the anisotropy parameters estimated by CHI, i.e., performing natural neighbor interpolation and anisotropy estimation using CHI in both exhaustive data sets and then plotting the non-parametric confidence regions and JPDFs. The results are shown

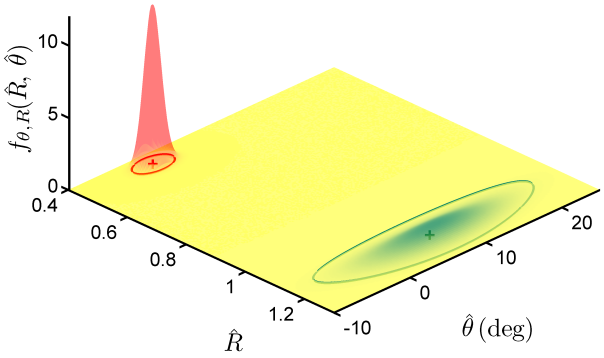


Fig. 9: Joint probability density functions (surfaces) and confidence 95% regions (solid contours) for the radioactivity dose rate data sets: normal data set (bottom right) and emergency data set (far left).

in Fig. 9. There is no overlap of the two density functions, and the contours corresponding to the 95% confidence regions are non-intersecting, which suggests a statistically significant difference of the anisotropy parameters between the background and the emergency data. Since the computation of the anisotropy estimates is very fast, our method provides an easy indicator of significant physical change in a system.

VII. Discussion and Conclusions

This work focuses on the estimation of anisotropy by means of the Covariance Hessian Identity (CHI) in two-dimensional digital data that are either scattered or supported on a grid. We derive explicit expressions for the joint PDF of anisotropy statistics, equation (18), and for the corresponding anisotropy confidence regions at any confidence level, equation (19). The main assumptions used are (i) that the data are drawn from a jointly Gaussian, stationary and differentiable random field (ii) that the covariance function is short-ranged. We also derive a non-parametric approximation for the joint PDF of the anisotropy statistics, which can be used if the covariance function is unknown *a priori*, or to avoid numerical calculations required for estimating the covariance. The non-parametric approximation is given by (23). The corresponding equation for the non-parametric approximation of the confidence region is given by (25). Practical application of the results of this research requires the estimation of anisotropy statistics using CHI (or other methods). Accurate estimation based on CHI requires a large sample size, $N \gg 1$ and a sample domain that is large with respect to the correlation area. The latter may be difficult to satisfy for data with large anisotropy ($R \gg 1$ or $R \ll 1$). In such cases, the CHI estimate tends to underestimate the actual anisotropy. We illustrate the application of the joint PDF and the confidence regions with simulated and real data. The results of this research can be used to identify significant deviations in anisotropy between data sets, e.g., due to structural differences or major changes in the underlying physical process. The computational cost is minimal, since the corresponding expressions are analytical. The major computational cost is the estimation of partial derivatives of the random field; the user can freely choose a more accurate field derivatives estimator, at the cost of additional computational time. Straightforward extension of this work is possible for jointly lognormal data along the lines of [11].

Appendix A
Jacobi's Theorems

We use Jabobi's theorems [18] to determine the PDF under the change-of-variables transformations $\vec{\mathbb{Q}} \to \hat{\mathbf{q}} \to (\hat{R}, \hat{\theta})$.

Theorem A.1 (Jabobi's univariate theorem)**.** *Let Z be a continuous random variable and $Y = g(Z)$, where $g(Z)$ is a continuous function in Z. If $y = g(z)$ admits at most a countable number of roots $z_j = g_j^{-1}(y),\ j = 1, \ldots, r$, then*

$$f_Y = \sum_{j=1}^{r} f_Z\big(g_j^{-1}(y)\big) \left| \frac{dg_j^{-1}(y)}{dy} \right|.$$

In the case of a multivariate variable transformation, Jacobi's theorem becomes:

Theorem A.2 (Jabobi's multivariate theorem)**.** *Let* $\mathbf{Z}$ *and* $\mathbf{Y}$ *be two* n*-dimensional continuous random vectors with components* $(Z_1, \ldots, Z_n)$ *and* $(Y_1, \ldots, Y_n)$*, respectively. The transformation* $\mathbf{Y} = \mathbf{g}(\mathbf{Z})$ *represents the set of equations* $y_l = g_l(\mathbf{Z}),\ l = 1, \ldots, n$*. Assume that the functions* g_l *are continuous and possess continuous partial derivatives with respect to each of their arguments.*

(i) If the g_l *define one-to-one mappings, unique inverse functions* g_l^{-1} *such that* $\mathbf{Z} = \mathbf{g}^{-1}(\mathbf{Y})$ *exist. Then, the transformation of the JPDF* $f_{\mathbf{Z}}$ *to* $f_{\mathbf{Y}}$ *is accomplished by means of*

$$f_{\mathbf{Y}} = f_{\mathbf{Z}}\left(\mathbf{g}^{-1}(\mathbf{Y})\right)\ |\det(\mathbf{J})|,$$

where $\mathbf{J}$ *is the Jacobian of the transformation:*

$$\mathbf{J} = \frac{\partial(g_1^{-1}, \ldots, g_n^{-1})}{\partial(y_1, \ldots, y_n)}.$$

(ii) If $\mathbf{Y} = \mathbf{g}(\mathbf{Z})$ *admits at most a countable number of roots* $\mathbf{Z}_j = \mathbf{g}_j^{-1}(\mathbf{y})$*,* $j = 1, \ldots, r$*, then*

$$f_{\mathbf{Y}} = \sum_{j=1}^{r} f_{\mathbf{Z}}\left(\mathbf{g}_j^{-1}(\mathbf{y})\right)\ |\det(\mathbf{J}_j)|,$$

where $\mathbf{J}_j$ *is the the Jacobian corresponding to the* j*-th root, defined by*

$$\mathbf{J}_j = \frac{\partial(g_{j,1}^{-1}, \ldots, g_{j,n}^{-1})}{\partial(y_1, \ldots, y_n)}.$$

Theorem A.2 also applies if $\dim(\mathbf{Y}) = m < n$. Then, the m-dimensional vector $\mathbf{Y}$ is augmented by an $(n-m)$-dimensional vector $\mathbf{Z}' = \mathbf{h}(\mathbf{Z})$ where $\mathbf{h}(\cdot)$ is a simple function with continuous partial derivatives. The $n - m$ dummy variables in $\mathbf{Z}'$ are eliminated from the JPDF of $\mathbf{Y}$ by integration.

APPENDIX B
PDF OF GRADIENT PRODUCT TENSOR

Below, we obtain the PDF of the gradient components $X_i = \partial_i X(\mathbf{s})$ for $i = 1, 2$. Let us define by $f_{X_i}(\partial_i X = z)$ the PDF of the gradient component $\partial_i X$, and by $f_{X_{ij}}(X_{ij} = y)$ the PDF of the gradient product $\partial_i X(\mathbf{s})\partial_j X(\mathbf{s})$. The following theorem holds [15]:

Theorem B.1 (PDF of field gradient)**.** *For a Gaussian, differentiable and stationary SRF* $X(\mathbf{s})$*, the gradient component* $\partial_i X(\mathbf{s})$ *is a zero-mean Gaussian SRF with covariance function given by the following expression:*

$$\mathbb{E}\left[\partial_i X(\mathbf{s})\, \partial_j X(\mathbf{s}+\mathbf{r})\right] = -\frac{\partial^2 c_{\mathrm{xx}}(\mathbf{r})}{\partial r_i \partial r_j}. \tag{B-27}$$

In light of (3) and (B-27) the *variance* of $\partial_i X(\mathbf{s})$ is given by $\mathrm{Var}\left(\partial_i X(\mathbf{s})\right) = Q_{ii}$. Hence, the univariate PDF f_{X_i} is given by

$$f_{X_i}(\partial_i X = z) = \frac{1}{\sqrt{2\pi\, Q_{ii}}}\, e^{-z^2/2Q_{ii}}. \tag{B-28}$$

A. PDF of diagonal elements of gradient product tensor

Theorem B.2 (PDF of diagonal elements of gradient product tensor)**.** *Let* $X(\mathbf{s})$ *be a Gaussian, statistically stationary SRF that admits (in the mean square sense) all partial derivatives of second-order (at least). Then, the PDF of* X_{ii} *is given by the chi square* (χ^2_ν) *distribution with one degree of freedom* $(\nu = 1)$*:*

$$f_{X_{ii}}(X_{ii} = y) = \frac{e^{-\frac{y}{2Q_{ii}}}}{\sqrt{2\pi Q_{ii}\, y}}. \tag{B-29}$$

Proof: Let us define $Z = \partial_i X(\mathbf{s})$, $(i = 1, 2)$ and $Y = X_{ii}$. The equation $y = g(z) = z^2$ admits two real roots for $y > 0$, i.e., $y_{1,2} = \pm\sqrt{z}$, but no real roots for $y < 0$. Since $\frac{dg^{-1}}{dy} = \pm\frac{1}{2\sqrt{y}}$, by applying Theorem A.1, we obtain:

$$f_{X_{ii}}(X_{ii} = y) = \frac{1}{2\sqrt{y}} \left[f_{X_i}(\sqrt{y}) + f_{X_i}(-\sqrt{y}) \right], \; y > 0. \tag{B-30}$$

Equation (B-29) follows from the above and (B-28). The standard density of the χ^2_1 distribution is obtained from (B-29) via the replacement $y' = y/Q_{ii}$. ■

The mean and variance of X_{ii} are thus obtained by

$$\mathbb{E}\left[X_{ii}\right] = Q_{ii}, \tag{B-31}$$

$$\mathrm{Var}\left(X_{ii}\right) = 2(Q_{ii})^2. \tag{B-32}$$

B. PDF of non-diagonal elements of gradient product tensor

Theorem B.3 (PDF of non-diagonal elements of the gradient product tensor)**.** *Let $X(\mathbf{s})$ be an SRF that satisfies the conditions of Theorem B.2. Then, the PDF of $X_{12} = \partial_1 X(\mathbf{s})\, \partial_2 X(\mathbf{s})$ is given by:*

$$f_{X_{12}}(X_{12} = y) = \frac{1}{\pi\sqrt{\det(\mathbf{Q})}} \exp\left(y\, \frac{Q_{12}}{\det(\mathbf{Q})} \right) K_0\left(|y|\, \frac{\sqrt{Q_{11}\, Q_{22}}}{\det(\mathbf{Q})} \right), \tag{B-33}$$

where $\mathbf{Q}$ is the slope tensor, $\det(\mathbf{Q})$ is its determinant, and K_0 is the modified Bessel function of the second kind and order zero. Since $Q_{12} = Q_{21}$, it holds that $f_{X_{21}}(X_{21} = y) = f_{X_{12}}(X_{12} = y)$.

Proof: We define the random vectors $\mathbf{Z}^t = (\partial_1 X(\mathbf{s}), \partial_2 X(\mathbf{s})) = (X_1, X_2)$, $\mathbf{Y}^t = (X_1 X_2, X_2)$. The Jacobian determinant for the transformation $\mathbf{Z} \to \mathbf{Y}$ and its absolute value are given by

$$\det(\mathbf{J}) = \begin{vmatrix} y_2^{-1} & -y_1\, y_2^{-2} \\ 0 & 1 \end{vmatrix} = \frac{1}{y_2} \Rightarrow |\det(\mathbf{J})| = \frac{1}{|y_2|}.$$

Thus, according to Theorem A.2, $f_{X_{12}}$ is given by

$$f_{X_{12}}(X_{12} = y_1) = \int_{-\infty}^{\infty} dy_2\, f_{zz}\left(y_1/y_2, y_2\right) \frac{1}{|y_2|}, \tag{B-34}$$

where f_{zz} is the bivariate Gaussian PDF of the two-component fluctuation (z_1, z_2), given by

$$f_{zz}(z_1, z_2) = \frac{1}{2\pi\, \det(\mathbf{Q})}\, e^{-\frac{1}{2}\mathbf{Z}^t \mathbf{Q}^{-1} \mathbf{Z}}, \tag{B-35}$$

Then, the integration in (B-34) can be evaluated using the integral (3.471.9) [31, p. 368] and the transformation $y_2^2 \to x$, which lead to (B-33). ■

The mean of X_{12} is given by (2). The variance is obtained using the normality of the $\partial_1 X(\mathbf{s})$ and $\partial_2 X(\mathbf{s})$ distributions by applying the Isserlis-Wick moment factorization theorem [32], [33]:

$$\mathbb{E}\left[X_{12}\right] = Q_{12}, \tag{B-36}$$

$$\mathrm{Var}\left(X_{12}\right) = Q_{11}\, Q_{22} + Q_{12}^2. \tag{B-37}$$

For $|x| \to \infty$, $K_0(|x|) \approx \sqrt{\frac{\pi}{2|x|}} e^{-|x|}$ [24]. Hence, based on (B-33), $f_{X_{12}}$ decays for large $|X_{12}|$ as

$$f_{X_{12}} \approx \frac{1}{\sqrt{2\pi}(Q_{11}\, Q_{22})^{1/4}} \frac{1}{\sqrt{|X_{12}|}} \exp\left[\frac{-|X_{12}|}{\sqrt{Q_{11}\, Q_{22}} + \mathrm{sign}(X_{12})\, Q_{12}} \right], \tag{B-38}$$

where $\mathrm{sign}(x) = 1, x > 0 \;\wedge\; \mathrm{sign}(x) = -1, x < 0$.

Asymptotic convergence of $f_{X_{12}}$ to zero for $|X_{12}| \to \infty$ requires that the denominator of the exponent be positive. Q_{11} and Q_{22} are positive by definition. Regardless of the sign of Q_{12}, the positive definiteness of $\mathbf{Q}$ (see Theorem II.1), implies that $\sqrt{Q_{11}\, Q_{22}} > \pm Q_{12}$. Hence, the denominator is indeed positive. The asymptotic dependence of $f_{X_{12}}$ matches that of the χ^2_1 distribution.

APPENDIX C
PROOF OF LEMMA III.1

Proof: Using the definition (8) we obtain:

$$C_{ij,kl} = \mathrm{Cov}\left(\frac{1}{N}\sum_{n=1}^{N} X_{ij}(\mathbf{s}_n), \frac{1}{N}\sum_{m=1}^{N} X_{kl}(\mathbf{s}_m)\right) = \frac{1}{N^2}\sum_{n,m}\mathrm{Cov}\left(X_{ij}(\mathbf{s}_n), X_{kl}(\mathbf{s}_m)\right). \tag{C-39}$$

Due to the translation invariance of $X(\mathbf{s})$, the double series in (C-39) is reduced to a single series over all (N^2) lag vectors $\mathbf{r}_{nm} = \mathbf{s}_n - \mathbf{s}_m$ $(n, m = 1, \ldots, N)$, i.e.,

$$\begin{aligned} C_{ij,kl} &= \frac{1}{N^2}\sum_{\mathbf{r}_{nm}}\mathrm{Cov}\left(X_{ij}(\mathbf{s}_0), X_{kl}(\mathbf{s}_0+\mathbf{r}_{nm})\right) \\ &= \frac{1}{N}\mathrm{Cov}\left(X_{ij}(\mathbf{0}), X_{kl}(\mathbf{0})\right) + \frac{1}{N^2}\sum_{\mathbf{r}_{nm}\neq\mathbf{0}}\mathrm{Cov}\left(X_{ij}(\mathbf{0}), X_{kl}(\mathbf{r}_{nm})\right). \end{aligned} \tag{C-40}$$

Covariance of the gradient product tensor: Let $\mathbf{r}$ denote any lag vector (including $\mathbf{r} = \mathbf{0}$) between two sampling points. Based on the definition of the covariance function it follows that

$$\mathrm{Cov}\left(X_{ij}(\mathbf{0}), X_{kl}(\mathbf{r})\right) = \mathbb{E}\left[X_{ij}(\mathbf{0})X_{kl}(\mathbf{r})\right] - \mathbb{E}\left[X_{ij}(\mathbf{0})\right]\mathbb{E}\left[X_{kl}(\mathbf{r})\right]. \tag{C-41}$$

Note that $\mathbb{E}\left[X_{ij}(\mathbf{0})X_{kl}(\mathbf{r})\right] = \mathbb{E}\left[\partial_i X(\mathbf{0})\,\partial_j X(\mathbf{0})\,\partial_k X(\mathbf{r})\,\partial_l X(\mathbf{r})\right]$. According to Theorem B.1, the gradient fields are Gaussian SRFs. Hence, $\mathbb{E}\left[X_{ij}(\mathbf{0})\,X_{kl}(\mathbf{r})\right]$ can be calculated using the moment factorization property of multivariate normal distributions [32], [33]:

$$\begin{aligned} \mathbb{E}\left[X_{ij}(\mathbf{0})\,X_{kl}(\mathbf{r})\right] &= \mathbb{E}\left[\partial_i X(\mathbf{0})\,\partial_j X(\mathbf{0})\right]\mathbb{E}\left[\partial_k X(\mathbf{r})\,\partial_l X(\mathbf{r})\right] + \mathbb{E}\left[\partial_i X(\mathbf{0})\,\partial_k X(\mathbf{r})\right]\mathbb{E}\left[\partial_j X(\mathbf{0})\,\partial_l X(\mathbf{r})\right] \\ &+ \mathbb{E}\left[\partial_i X(\mathbf{0})\,\partial_l X(\mathbf{r})\right]\mathbb{E}\left[\partial_j X(\mathbf{0})\,\partial_k X(\mathbf{r})\right] = H_{ij}(\mathbf{0})\,H_{kl}(\mathbf{0}) + H_{ik}(\mathbf{r})\,H_{jl}(\mathbf{r}) + H_{il}(\mathbf{r})\,H_{jk}(\mathbf{r}). \end{aligned} \tag{C-42}$$

The last equality follows from the definition of the covariance function, Theorem B.1, and the definition (1). The second term on the right-hand side of (C-41) is simply

$$\mathbb{E}\left[X_{ij}(\mathbf{0})\right]\mathbb{E}\left[X_{kl}(\mathbf{r})\right] = \mathbb{E}\left[\partial_i X(\mathbf{0})\,\partial_j X(\mathbf{0})\right]\mathbb{E}\left[\partial_k X(\mathbf{r})\,\partial_l X(\mathbf{r})\right] = H_{ij}(\mathbf{0})\,H_{kl}(\mathbf{0}). \tag{C-43}$$

Thus, in light of (C-42) and (C-43), equation (C-41) becomes

$$\mathrm{Cov}\left(X_{ij}(\mathbf{0}), X_{kl}(\mathbf{r})\right) = H_{ik}(\mathbf{r})H_{jl}(\mathbf{r}) + H_{il}(\mathbf{r})H_{jk}(\mathbf{r}). \tag{C-44}$$

Using (C-44) in (C-40), and Theorem II.1 to express the zero-lag component of the covariance Hessian matrix, equation (10) is obtained.

■

APPENDIX D
PROOF OF THEOREM III.1

Proof: The $X_{ij}(\mathbf{s}_k)$ are stationary GSRFs by virtue of the stationarity of the GSRF $X(\mathbf{s})$. Hence, $\phi_{ijkl}(\mathbf{r}) := \mathrm{Cov}\left(X_{ij}(\mathbf{s}), X_{kl}(\mathbf{s}+\mathbf{r})\right) = \mathrm{Cov}\left(X_{ij}(\mathbf{0}), X_{kl}(\mathbf{r})\right)$. Using (C-44), $\phi_{ijkl}(\mathbf{r}) = H_{ik}(\mathbf{r})H_{jl}(\mathbf{r}) + H_{il}(\mathbf{r})H_{jk}(\mathbf{r})$. The range of the GSRF $X_{ij}(\mathbf{s}_k)$ is determined by the integral

$$V_c = \max_{i,j,k,l}\left(\frac{1}{\phi_{ijkl}(\mathbf{0})}\int d\mathbf{r}\phi_{ijkl}(\mathbf{r}),\ \Big|\ \phi_{ijkl}(\mathbf{0}) \neq 0\right).$$

Based on (C-44), $\phi_{ijkl}(\mathbf{0}) = Q_{ij}\,Q_{kl} + Q_{il}\,Q_{jk}$ and thus $\phi_{ijkl}(\mathbf{0})$ has a finite value for finite correlation lengths.

To calculate $\int_{\mathcal{D}} d\mathbf{r}\phi_{ijkl}(\mathbf{r})$, we assume that $|\mathcal{D}| \to \infty$ and express the integral in terms of the Fourier transform of $c_{\mathrm{xx}}(\mathbf{r})$. Any permissible covariance function $c_{\mathrm{xx}}(\mathbf{r})$ admits the following pair of transformations, where $\tilde{C}_{\mathrm{xx}}(\mathbf{k})$ is the *covariance spectral density*:

$$\begin{aligned} c_{\mathrm{xx}}(\mathbf{r}) &= \frac{1}{(2\pi)^2}\int d\mathbf{k}e^{\jmath\mathbf{k}\cdot\mathbf{r}}\tilde{C}_{\mathrm{xx}}(\mathbf{k}), \\ \tilde{C}_{\mathrm{xx}}(\mathbf{k}) &= \int d\mathbf{r}e^{-\jmath\mathbf{k}\cdot\mathbf{r}}\,c_{\mathrm{xx}}(\mathbf{r}). \end{aligned}$$

Based on the above, we obtain $H_{ij}(\mathbf{r}) = \frac{1}{(2\pi)^2}\int d\mathbf{k} k_i k_j\, e^{\jmath \mathbf{k}\cdot\mathbf{r}} \tilde{C}_{\rm xx}(\mathbf{k})$, and thus

$$\int d\mathbf{r}\phi_{ijkl}(\mathbf{r}) = \frac{1}{(2\pi)^2}\int d\mathbf{k} k_i k_j k_k k_l [\tilde{C}_{\rm xx}(\mathbf{k})]^2.$$

In the above, $\jmath = \sqrt{-1}$, $\mathbf{k}\cdot\mathbf{r} = k_1 r_1 + k_2 r_2$ is the inner vector product, and $\int d\mathbf{k} = \int_{-\infty}^{\infty} dk_1 \int_{-\infty}^{\infty} dk_2$ or $\int d\mathbf{k} = \int_0^{\infty} k\, dk \int_0^{2\pi} d\phi$ in polar coordinates. Existence of this integral requires integrability of $\tilde{C}_{\rm xx}(\mathbf{k})$ at $k=0$ and at $k\to\infty$. Since $c_{\rm xx}(\mathbf{r})$ is short-ranged, the integral $\int d\mathbf{r}\, c_{\rm xx}(\mathbf{r}) = \tilde{C}_{\rm xx}(\mathbf{0})$ is finite, and thus the limit $k=0$ is well-behaved. At the limit $k\to\infty$, the integral converges (using polar coordinates) if $[\tilde{C}_{\rm xx}(\mathbf{k})]^2$ falls off faster than $k^{-6-2\epsilon}$, where $\epsilon>0$. This ensures that the covariance $\phi_{ijkl}(\mathbf{r})$ is short-ranged.

Thus, Theorem III.1 applies to the vector random variable $\mathbf{Z}_k = ((X_{11}(\mathbf{s}_k), X_{22}(\mathbf{s}_k), X_{12}(\mathbf{s}_k))^t$ leading to (11). ■

APPENDIX E
PROOF OF LEMMA III.3

Proof: We use Theorem A.2 with $\mathbf{Z}\to\vec{\mathbb{Q}}$, $\mathbf{Y}\to(\hat{q}_d,\hat{q}_o)^t$. Since $\dim(\mathbf{Y}) = 2 < \dim(\mathbf{Z}) = 3$, we append to $\mathbf{Y}$ the dummy variable $u = \hat{Q}_{11} \geq 0$. Using definitions (4) and (5), the absolute value of the Jacobian determinant for the transformation $(\hat{Q}_{11},\hat{Q}_{22},\hat{Q}_{12}) \to (\hat{q}_d,\hat{q}_o,u)$ is

$$\mathbf{J}_{\mathbf{q}} = \frac{\partial(\hat{Q}_{11},\hat{Q}_{22},\hat{Q}_{12})}{\partial(\hat{q}_d,\hat{q}_o,u)} \Rightarrow |\det(\mathbf{J}_{\mathbf{q}})| = u^2.$$

The dummy variable u is integrated according to Theorem A.2, leading to

$$f_{\mathbf{q}}(\hat{\mathbf{q}};\mathbf{m}_{\mathbb{Q}},\mathbf{C}_{\vec{\mathbb{Q}}}) = \int_0^{\infty} f_{\vec{\mathbb{Q}}}(u,\hat{q}_o u,\hat{q}_d u;\mathbf{m}_{\mathbb{Q}},\mathbf{C}_{\vec{\mathbb{Q}}})\, u^2\, du. \tag{E-45}$$

In terms of $\hat{q}_d$ and $\hat{q}_o$, the exponent of the trivariate PDF (11) is transformed as follows

$$(\vec{\mathbb{Q}}-\mathbf{m}_{\mathbb{Q}})^t\, \mathbf{C}_{\vec{\mathbb{Q}}}^{-1}\, (\vec{\mathbb{Q}}-\mathbf{m}_{\mathbb{Q}}) = A(\hat{\mathbf{q}},\mathbf{C}_{\vec{\mathbb{Q}}})\, u^2 + B(\hat{\mathbf{q}},\mathbf{m}_{\mathbb{Q}},\mathbf{C}_{\vec{\mathbb{Q}}})\, u + C(\mathbf{m}_{\mathbb{Q}},\mathbf{C}_{\vec{\mathbb{Q}}}). \tag{E-46}$$

By virtue of the above, the integral (E-45) is expressed as follows (suppressing the dependence of A,B,C,K for brevity):

$$f_{\mathbf{q}}(\hat{\mathbf{q}};\mathbf{m}_{\mathbb{Q}},\mathbf{C}_{\vec{\mathbb{Q}}}) = K\int_0^{\infty} u^2\, e^{-\frac{1}{2}(Au^2+Bu+C)}\, du.$$

According to (14a), $A>0$ because $\mathbf{C}_{\vec{\mathbb{Q}}}$ is a covariance matrix, and hence $\mathbf{C}_{\vec{\mathbb{Q}}}$ as well as $\mathbf{C}_{\vec{\mathbb{Q}}}^{-1}$ are positive definite[5]. Thus, the Gaussian integral above exists and its value is given by (13).

■

APPENDIX F
PROOF OF THEOREM III.2

Proof: Equation (18) follows from the transformation $(\hat{q}_d,\hat{q}_o)\to(\hat{R},\hat{\theta})$ with Jacobian matrix $\mathbf{J}_{\theta,R}$. According to A.2 the transformed PDF is given by $f_{\theta,R}(\hat{R},\hat{\theta};\mathbf{m}_{\mathbb{Q}},\mathbf{C}_{\vec{\mathbb{Q}}}) = f_{\mathbf{q}}(\hat{\mathbf{q}};\mathbf{m}_{\mathbb{Q}},\mathbf{C}_{\vec{\mathbb{Q}}})\ |\det(\mathbf{J}_{\theta,R})|$ where $\det(\mathbf{J}_{\theta,R})$ is given by

$$\det(\mathbf{J}_{\theta,R}) = \begin{vmatrix} \frac{\partial\hat{q}_d}{\partial\hat{\theta}} & \frac{\partial\hat{q}_d}{\partial\hat{R}} \\ \frac{\partial\hat{q}_o}{\partial\hat{\theta}} & \frac{\partial\hat{q}_o}{\partial\hat{R}} \end{vmatrix} = \frac{2\hat{R}\left(\hat{R}^2-1\right)}{\left(\hat{R}^2\cos^2\hat{\theta}+\sin^2\hat{\theta}\right)^3}. \tag{F-47}$$

Under the restriction of the parameter space to $R\in[0,\infty)$ and $\theta\in[-\pi/4,\pi/4)$, or equivalently $R\in[1,\infty)$ and $\theta\in[-\pi/2,\pi/2)$, the transformation $(\hat{q}_d,\hat{q}_o)\to(\hat{R},\hat{\theta})$ is one-to-one with a Jacobian determinant given by (F-47). Finally, based on Theorems A.2 and III.3, $f_{\theta,R}(\hat{R},\hat{\theta})$ is given by (18). ■

[5]A square $p\times p$ matrix $\mathbf{M}$ is positive-definite, denoted by $\mathbf{M}>0$, if $\mathbf{x}^t\,\mathbf{M}\,\mathbf{x}>0\ \forall\ p\times 1$ vectors $\mathbf{x}\neq\mathbf{0}$; then, $\mathbf{M}^{-1}>0$.

REFERENCES

[1] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Boston, MA: MIT Press, 2006.

[2] X. Jiang, "On orientation and anisotropy estimation for online fingerprint authentication," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 4038 – 4049, 2005.

[3] K. Okada, D. Comaniciu, and A. Krishnan, "Robust anisotropic Gaussian fitting for volumetric characterization of pulmonary nodules in multislice CT," *IEEE Transactions on Medical Electronics*, vol. 24, no. 3, pp. 409–423, 2005.

[4] L. Feng, I. Hotz, B. Hamann, and K. I. Joy, "Anisotropic noise samples," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 2, pp. 342–354, 2008.

[5] S. C. Olhede, "Localisation of geometric anisotropy," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 2133–2137, 2008.

[6] D. l. Bihan, J. F. Mangin, C. Poupon, N. M. C. A. Clark, S. Pappata, and H. Chabriat, "Diffusion tensor imaging: concepts and applications," *Journal of Magnetic Resonance Imaging*, vol. 13, no. 4, pp. 534–546, 2001.

[7] Y. Xu and J. Choi, "Mobile sensor networks for learning anisotropic Gaussian processes," in *American Control Conference, 2009. ACC '09*, june 2009, pp. 5049 –5054.

[8] F. Richard and H. Bierme, "Statistical tests of anisotropy for fractional Brownian textures. Application to full-field digital mammography," *Journal of Mathematical Imaging and Vision*, vol. 36, pp. 227–240, 2010.

[9] L. Wang and C. A. Leckie, "Improved Gaussian process classification via feature space rotation," *Neurocomputing*, vol. 83, pp. 89–97, April 2012.

[10] D. Hristopulos, "New anisotropic covariance models and estimation of anisotropic parameters based on the covariance tensor identity," *Stochastic Environmental Research and Risk Assessment*, vol. 16, no. 1, pp. 43–62, 2002.

[11] A. Chorti and D. T. Hristopulos, "Nonparametric identification of anisotropic (elliptic) correlations in spatially distributed data sets," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4738–4751, Oct. 2008.

[12] S. I. Ranganathan, M. Ostoja-Starzewski, and M. Ferrari, "Quantifying the anisotropy in biological materials," *Journal of Applied Mechanics*, vol. 78, no. 6, p. 064501, 2011.

[13] E. Pebesma, D. Cornford, G. Dubois, G. Heuvelink, D. Hristopoulos, J. Pilz, U. Stoehlker, G. Morin, and J. Skoien, "Intamap: The design and implementation of an interoperable automated interpolation web service," *Computers and Geosciences*, vol. 37, pp. 343–352, 2011.

[14] A. M. Schmidt and A. O'Hagan, "Bayesian inference for non-stationary spatial covariance structure via spatial deformations," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 65, no. 3, pp. 743–758, 2003.

[15] A. M. Yaglom, *Correlation Theory of Stationary and Related Random Functions I*. New York: Springer Verlag, 1987.

[16] R. J. Adler, *The Geometry of Random Fields*, 1st ed. New York: Wiley, 1981.

[17] P. Swerling, "Statistical properties of the contours of random surfaces," *IRE Transactions on Information Theory*, vol. IT-8, pp. 315–321, Jul. 1962.

[18] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 2002.

[19] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*. Reading, MA: Addison Wesley, 1954.

[20] P. Levy, *Théorie de l' Addition des Variables Aléatoires*. Paris: Gauthier Villars, 1954.

[21] W. Feller, *An introduction to Probability Theory*. New York: Wiley, 1971.

[22] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd ed. New York: Wiley, 1984.

[23] J.-P. Bouchaud and A. Georges, "Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications," *Physics Reports*, vol. 195, pp. 127–293, Nov. 1990.

[24] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, 1st ed., ser. Dover books on mathematics. Dover Publications, June 1970.

[25] M. Siotani, "Tolerance regions for a multivariate normal population," *Annals of the Institute of Statistical Mathematics*, vol. 16, no. 1, pp. 135–153, 1964.

[26] E. Pardo-Igúzquiza and M. Chica-Olmo, "The Fourier integral method: an efficient spectral method for simulation of random fields," *Mathematical Geology*, vol. 25, no. 2, pp. 177–217, 1993.

[27] C. Lantuéjoul, *Geostatistical Simulation: Models and Algorithms*. New York: Springer, 2002.

[28] D. Hristopulos and M. Žukovič, "Relationships between correlation lengths and integral scales for covariance models with more than two parameters," *Stochastic Environmental Research and Risk Assessment*, vol. 25, no. 1, pp. 11–19, 2011.

[29] P. Fisher, H. Ledoux, and C. Gold, "An efficient natural neighbour interpolation algorithm for geoscientific modelling," in *Developments in Spatial Data Handling*. Springer Berlin Heidelberg, 2005, pp. 97–108.

[30] G. Dubois and S. Galmarini, "Spatial interpolation comparison (SIC) 2004: introduction to the exercise and overview of results," in *Automatic Mapping Algorithms for Routine and Emergency Monitoring*, G. Dubois, Ed. Luxembourg, European Communities: Office for Official Publications of the European Communities, 2006, vol. EUR-21595-EN, pp. 7–18.

[31] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. New York: Elsevier, 2007.

[32] L. Isserlis, "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," *Biometrika*, vol. 12, no. 1-2, pp. 134–139, 1918.

[33] G. C. Wick, "The evaluation of the collision matrix," *Physical Review*, vol. 80, no. 2, pp. 268–272, 1950.